

Escola de Gestão do Porto

**A construção de um sistema de Armazenamento de
Dados no âmbito do sistema GIST98/EUROBUS**

João Pedro Carvalho Leal Mendes Moreira

Licenciado em Matemática e Ciências de Computação pela
Universidade do Minho

Dissertação submetida para satisfação parcial dos requisitos do grau de mestre em
Gestão de Empresas

Dissertação realizada sob supervisão de
Professor Doutor Jorge Rui Guimarães Freire de Sousa,
do Departamento de Engenharia Mecânica e Gestão Industrial
da Faculdade de Engenharia da Universidade do Porto

Porto, Novembro de 2000

Palavras chave - Key words

Português

Armazém de Dados

Armazém de Dados Departamental

Armazenamento de Dados

Bases de Dados multi-dimensionais

Processamento analítico

Sistemas de Apoio à Decisão

Transportes

Transportes colectivos rodoviários de passageiros

English

Analytical Processing

Data Mart

Data Warehouse

Data Warehousing

Decision Support Systems

Mass Transit

Multidimensional Databases

OLAP

Transportation

Resumo

Neste trabalho descreve-se o desenvolvimento do Módulo Quadro de Indicadores Operacionais (Módulo QIO) como um sistema de Armazenamento de Dados. Este módulo tem por objectivo disponibilizar informação de planeamento e controlo aos agentes de decisão de empresas de transportes colectivos rodoviários de passageiros. Faz parte integrante do sistema GIST98/EUROBUS – um Sistema de Apoio à Decisão para a área do planeamento.

Começam-se por definir os conceitos fundamentais sobre sistemas de Armazenamento de Dados e apresentar métodos adequados para o desenvolvimento deste tipo de sistemas. De seguida apresenta-se o caso de estudo (a construção do Módulo QIO no âmbito do sistema GIST98/EUROBUS), descreve-se o método de desenvolvimento a utilizar na construção do Módulo QIO e, por último, descreve-se a aplicação do método.

O trabalho apresentado foi realizado em colaboração com cinco empresas que utilizam ou vão utilizar num futuro próximo o sistema GIST98/EUROBUS. Estas empresas operam diariamente com cerca de 6000 viaturas correspondendo a mais de metade do mercado de transportes colectivos rodoviários de passageiros em Portugal, incluindo Madeira e Açores.

Abstract

This work describes the development of the Performance Indicators Board Module (PIB Module) as a Data Warehousing system. This module aims to give planning and control information to support the decision making process at mass transit companies. It is part of the GIST98/EUROBUS system – a Decision Support System to the planning area.

Firstly are defined the main concepts about Data Warehousing systems and presented methods to develop this kind of systems. Next is presented the study case (the PIB Module construction as a module of the GIST98/EUROBUS system), described the development method to use on the study case and, finally, it is described the method application.

This work was done with the support of five companies that are using or will use soon the GIST98/EUROBUS system. Those companies operate daily about 6000 vehicles, corresponding to more than half of the road public transport market in Portugal, including Madeira and Azores.

Agradecimentos

Em primeiro lugar, quero agradecer ao meu supervisor, Professor Jorge Freire de Sousa, todo o apoio e disponibilidade, assim como as sugestões que deu e o envolvimento que a empresa STCP teve ao longo deste trabalho, em grande parte pelo seu empenho pessoal.

Ao Professor João Falcão e Cunha como responsável pelo sistema GIST98/EUROBUS e, ainda, pelos livros arrançados e sugestões apresentadas.

A todas as empresas do consórcio GIST, nomeadamente CARRIS, STCP, Grupo Barraqueiro, Vimeca e Horários do Funchal toda a disponibilidade e empenho, em particular na fase de estudo dos requisitos do negócio.

Ao GEIN e à Escola de Gestão do Porto as condições de trabalho oferecidas.

À Novabase Porto nas pessoas do Engº João Ranito, Engº Luís Gama e Dr. Rui Almeida os esclarecimentos prestados e a bibliografia disponibilizada.

Ao Professor Orlando Belo da Universidade do Minho todos os esclarecimentos prestados.

Ao Zé Luís pela revisão do texto e ao Barros Basto todo o apoio dado.

A toda a equipa do sistema GIST98/EUROBUS. Num trabalho como este, a disponibilidade das pessoas envolvidas nos diferentes módulos foi essencial para o esclarecimento das minhas (muitas) dúvidas. Isso só foi possível dada a grande amizade existente entre todos os elementos deste grupo fantástico.

Ao Dinis, companheiro de armas já lá vão muitos anos, para além da ajuda na revisão do texto e como elemento da equipa do sistema GIST98/EUROBUS, o ter aguentado as minhas fases de maior irritação quando as coisas corriam pior e aturava os meus desabafos cá com uma paciência ... de amigo, naturalmente.

De modo muito especial ao João Castro que foi o meu verdadeiro braço direito, em particular no desenvolvimento do processo de aquisição de dados. Para além disso, as muitas discussões construtivas mantidas foram de extrema importância para questionar e consolidar ideias.

Este trabalho foi apoiado pelo Programa Praxis XXI - Medida 3.1b) "Investigação em Consórcio", Ref. P015-P31B-09/96-EUROBUS.

Índice

1	INTRODUÇÃO.....	1
1.1	ORGANIZAÇÃO DA DISSERTAÇÃO	2
1.2	ALGUMAS CONSIDERAÇÕES SOBRE A ELABORAÇÃO DA DISSERTAÇÃO	2
2	CONCEITOS FUNDAMENTAIS	4
2.1	DEFINIÇÕES	4
2.1.1	<i>Sistema de Apoio à Decisão.....</i>	<i>4</i>
2.1.2	<i>Armazém de Dados (Data Warehouse).....</i>	<i>4</i>
2.1.3	<i>Armazenamento de Dados (Data Warehousing).....</i>	<i>6</i>
2.1.4	<i>Armazém de Dados Departamental (Data Mart).....</i>	<i>6</i>
2.1.5	<i>Granulosidade.....</i>	<i>7</i>
2.1.6	<i>Multi-dimensionalidade: conceitos associados</i>	<i>7</i>
2.1.7	<i>OLAP - processamento analítico por controlo directo (On-Line Analytical Processing)</i>	<i>9</i>
2.1.8	<i>Servidor OLAP.....</i>	<i>10</i>
2.2	ARQUITECTURA DO ARMAZENAMENTO DE DADOS	11
2.2.1	<i>Aquisição dos dados</i>	<i>12</i>
2.2.2	<i>Administração das estruturas de informação</i>	<i>13</i>
2.2.3	<i>Ferramentas para o apoio à decisão</i>	<i>14</i>
3	MÉTODOS DE DESENVOLVIMENTO	17
3.1	FACTORES CRÍTICOS DE SUCESSO.....	17
3.2	APRESENTAÇÃO DOS MÉTODOS	19
3.2.1	<i>Método Hadden – Kelly [Hadden & Company, 1999]</i>	<i>19</i>
3.2.2	<i>Método SDW [Ebel, 1998], [Gardner, 1998]</i>	<i>21</i>
3.2.3	<i>Método da DataWing Consulting Services, LLC [Welch, 1998]</i>	<i>25</i>
3.3	OS MÉTODOS APRESENTADOS FACE AOS FACTORES CRÍTICOS DE SUCESSO	27
4	APRESENTAÇÃO DO CASO DE ESTUDO.....	29
4.1	O CONSÓRCIO GIST	29
4.2	O SISTEMA GIST98/EUROBUS [ICAT & INEGI, 1997]	30
4.2.1	<i>A definição da rede de transportes: o Módulo Rede.....</i>	<i>31</i>
4.2.2	<i>A definição das linhas: o Módulo Megalinhas.....</i>	<i>32</i>
4.2.3	<i>A definição das viagens e do serviço das viaturas: o Módulo Viagens e Viaturas</i>	<i>34</i>
4.2.4	<i>A definição dos serviços a realizar e a sua afectação aos tripulantes: os módulos de Serviços e de Escalamento.....</i>	<i>35</i>

4.2.5	<i>A disponibilização de informação ao público: o Módulo de Informação ao Público.....</i>	41
4.2.6	<i>A disponibilização de informação aos agentes de decisão: o Módulo QIO.....</i>	42
4.2.7	<i>O conceito de versão activa.....</i>	43
4.2.8	<i>As Bases de Dados do sistema</i>	43
5	O MÓDULO QIO FACE AOS SISTEMAS DE ARMAZENAMENTO DE DADOS	44
5.1	<i>A CONSTRUÇÃO DE UM ARMAZÉM DE DADOS NO ÂMBITO DO SISTEMA GIST98/EUROBUS</i>	44
5.2	<i>A ESPECIFICIDADE DO MÉTODO DE DESENVOLVIMENTO PARA O CASO EM ESTUDO</i>	45
6	APLICAÇÃO DO MÉTODO DE DESENVOLVIMENTO.....	48
6.1	<i>ESTUDO DOS REQUISITOS DO NEGÓCIO</i>	48
6.2	<i>ESTUDO DAS FONTES DE DADOS.....</i>	56
6.3	<i>DESENHO DA ARQUITECTURA DO ARMAZENAMENTO DE DADOS.....</i>	60
6.3.1	<i>Estudo da dimensão da Base de Dados</i>	60
6.3.2	<i>Seleção das ferramentas OLAP a analisar</i>	64
6.3.3	<i>Avaliação das ferramentas OLAP.....</i>	65
6.3.4	<i>Definição da arquitectura geral do sistema.....</i>	67
6.4	<i>INSTALAÇÃO DA INFRA-ESTRUTURA DE ARMAZENAMENTO DE DADOS</i>	68
6.5	<i>DESENHO E CONSTRUÇÃO DA BASE DE DADOS E DO PROCESSO DE AQUISIÇÃO DOS DADOS.....</i>	68
6.5.1	<i>O modelo dimensional</i>	68
6.5.2	<i>Desenho e construção da Base de Dados [Microsoft Corporation, 1998].....</i>	69
6.5.3	<i>O processo de aquisição de dados</i>	72
6.6	<i>INSTALAÇÃO E CONFIGURAÇÃO DAS FERRAMENTAS PARA O APOIO À DECISÃO</i>	81
6.6.1	<i>A configuração da ferramenta OLAP: os serviços OLAP do SQL Server 7.....</i>	81
6.6.2	<i>Exemplos de ferramentas para o utilizador final.....</i>	85
6.7	<i>A ETAPA DE IMPLEMENTAÇÃO</i>	90
7	CONCLUSÕES.....	92
7.1	<i>AVALIAÇÃO DO TRABALHO DESENVOLVIDO</i>	92
7.2	<i>POSSÍVEIS DESENVOLVIMENTOS DO MÓDULO QIO.....</i>	93
7.2.1	<i>Na perspectiva de evolução do sistema GIST 98/EUROBUS</i>	93
7.2.2	<i>Na perspectiva das empresas do consórcio GIST.....</i>	93
7.3	<i>ARMAZÉM DE DADOS OU ARMAZÉM DE DADOS DEPARTAMENTAL?</i>	94
	ANEXO A: INDICADORES PROPOSTOS PELAS EMPRESAS.....	96
	<i>STCP.....</i>	96
	<i>GRUPO BARRAQUEIRO</i>	100
	<i>Rodoviária de Lisboa.....</i>	100
	<i>EVA Transportes.....</i>	101
	<i>HORÁRIOS DO FUNCHAL.....</i>	102

ANEXO B: ESQUEMAS RELACIONAIS DA BASE DE DADOS DE PLANEAMENTO	104
MÓDULO REDE.....	104
MÓDULO MEGALINHAS	105
MÓDULO VIAGENS E VIATURAS.....	106
MÓDULO SERVIÇOS.....	107
MÓDULO ADMINISTRAÇÃO	109
MÓDULO GESTÃO DE PESSOAL TRIPULANTE.....	110
MÓDULO PRÉ-ESCALAMENTO	111
MÓDULOS GERAÇÃO DE ESCALAS E PÓS-ESCALAMENTO	112
BIBLIOGRAFIA E REFERÊNCIAS	113

Índice de figuras

FIGURA 1.1: O MERCADO MUNDIAL DE FERRAMENTAS OLAP EM MILHARES DE MILHÕES DE US\$ ([PENDSE & CREETH, 2000A])	1
FIGURA 2.1: ARQUITECTURA DE REFERÊNCIA DO ARMAZENAMENTO DE DADOS (ADAPTADO DE [RANITO, 1999])	11
FIGURA 2.2: ARQUITECTURA <i>ROLAP</i>	15
FIGURA 2.3: ARQUITECTURA <i>MOLAP</i>	15
FIGURA 2.4: ARQUITECTURA <i>HOLAP</i>	15
FIGURA 3.1: MÉTODO <i>HADDEN-KELLY</i> ([HADDEN & COMPANY, 1999])	20
FIGURA 3.2: MÉTODO SDW (TRADUZIDO DE [EBEL, 1998])	22
FIGURA 3.3: MÉTODO SDW DE GESTÃO (TRADUZIDO DE [EBEL, 1998])	22
FIGURA 3.4: MÉTODO SDW DETALHADO (TRADUZIDO DE [GARDNER, 1998])	23
FIGURA 3.5: MÉTODO DA DATAWING CONSULTING SERVICES, LLC DETALHADO (TRADUZIDO DE [WELCH, 1998])	26
FIGURA 3.6: COMPARAÇÃO ENTRE OS MÉTODOS DE DESENVOLVIMENTO	27
FIGURA 3.7: MÉTODO DE DESENVOLVIMENTO GENÉRICO	28
FIGURA 4.1: O CONSÓRCIO GIST	30
FIGURA 4.2: ARQUITECTURA DE REFERÊNCIA DO SISTEMA GIST98/EUROBUS	30
FIGURA 4.3: MÓDULO REDE	31
FIGURA 4.4: MÓDULO MEGALINHAS	33
FIGURA 4.5: MÓDULO VIAGENS E VIATURAS	35
FIGURA 4.6: MÓDULO SERVIÇOS	37
FIGURA 4.7: MÓDULO GESTÃO DE PESSOAL TRIPULANTE	37
FIGURA 4.8: MÓDULO PRÉ-ESCALAMENTO	39
FIGURA 4.9: MÓDULO GERAÇÃO DE ESCALAS	41
FIGURA 4.10: COMPONENTE INFOBUS DO MÓDULO DE INFORMAÇÃO AO PÚBLICO	42
FIGURA 5.1: MÉTODO DE DESENVOLVIMENTO DO MÓDULO QIO	45
FIGURA 6.1: ESQUEMA EM ESTRELA (TRADUZIDO DE [RADEN, 1995/6])	61
FIGURA 6.2: ARQUITECTURA DO MÓDULO QIO (ADAPTADO DE [RANITO, 1999])	67
FIGURA 6.3: ESQUEMA EM CRISTAL DE GELO	69
FIGURA 6.4: ESQUEMA RELACIONAL PARA O CUBO ‘REDE COMERCIAL’	70
FIGURA 6.5: ESQUEMA RELACIONAL PARA O CUBO ‘SERVIÇOS DAS VIATURAS’	71
FIGURA 6.6: ESQUEMA RELACIONAL PARA O CUBO ‘PESSOAL TRIPULANTE’	71
FIGURA 6.7: ESQUEMA RELACIONAL PARA AS TABELAS SYS_	72
FIGURA 6.8: EXEMPLO DE UMA TABELA MAP_	72
FIGURA 6.9: ESQUEMA RELACIONAL PARA AS TABELAS AUX_	73
FIGURA 6.10: ESQUEMA RELACIONAL PARA AS TABELAS TEMPQIO_	73

FIGURA 6.11: EXEMPLO DE UMA JANELA DE AVISO	74
FIGURA 6.12: JANELA PARA A AUTOMATIZAÇÃO DE UM PROCESSO	78
FIGURA 6.13: LEGENDA DO ESQUEMA DE MIGRAÇÃO DE DADOS	78
FIGURA 6.14: ESQUEMA DE MIGRAÇÃO DOS DADOS	79
FIGURA 6.15: SELECÇÃO DOS NÍVEIS DE UMA DIMENSÃO TEMPO	81
FIGURA 6.16: ESQUEMA RELACIONAL USADO NA DEFINIÇÃO DE UMA DIMENSÃO COM VÁRIAS TABELAS	82
FIGURA 6.17: JANELA PARA A CONSTRUÇÃO DE INDICADORES CALCULADOS	83
FIGURA 6.18: JANELA PARA A ESCOLHA DA FORMA DE ARMAZENAMENTO DOS DADOS	84
FIGURA 6.19: JANELA PARA A DEFINIÇÃO DAS OPÇÕES DE AGREGAÇÃO.....	85
FIGURA 6.20: <i>MICROSOFT EXCEL 2000</i> - JANELA PARA A DEFINIÇÃO DE UMA MATRIZ.....	86
FIGURA 6.21: <i>MICROSOFT EXCEL 2000</i> - MATRIZ COM DADOS DO CUBO ‘SERVIÇOS DAS VIATURAS’	86
FIGURA 6.22: <i>MICROSOFT EXCEL 2000</i> - GRÁFICO COM O SERVIÇO DAS VIATURAS AO LONGO DO DIA	87
FIGURA 6.23: <i>KNOSYS PROCLARITY 2.0</i> – MATRIZ	88
FIGURA 6.24: <i>KNOSYS PROCLARITY 2.0</i> – GRÁFICO DE BARRAS	88
FIGURA 6.25: <i>KNOSYS PROCLARITY 2.0</i> – DECOMPOSIÇÃO EM ÁRVORE.....	89
FIGURA 6.26: <i>KNOSYS PROCLARITY 2.0</i> – ANÁLISE DE CORRELAÇÃO	89

Índice de tabelas

TABELA 2.1: COMPARAÇÃO ENTRE OS DADOS OPERACIONAIS E OS DADOS DE INFORMAÇÃO (TRADUZIDO DE [INMON, 1996]).....	5
TABELA 2.2: COMPARAÇÃO ENTRE OS ARMAZÉNS DE DADOS DEPARTAMENTAIS E OS ARMAZÉNS DE DADOS (TRADUZIDO DE [WATTERSON, 1997])	7
TABELA 2.3: CARACTERÍSTICAS DE QUALIDADE (TRADUZIDO DE [HUFFORD, 1996]).....	12
TABELA 6.1: 1ª VERSÃO DAS DIMENSÕES.....	49
TABELA 6.2: 1ª VERSÃO DOS INDICADORES DE EXPLORAÇÃO	50
TABELA 6.3: 1ª VERSÃO DOS INDICADORES SOBRE MOTORISTAS	50
TABELA 6.4: VERSÃO FINAL DAS DIMENSÕES.....	51
TABELA 6.5: RELAÇÃO DAS DIMENSÕES POR CUBO	54
TABELA 6.6: VERSÃO FINAL DOS INDICADORES	54
TABELA 6.7: MÓDULOS DE ORIGEM DOS DADOS PARA A BASE DE DADOS DE ANÁLISE E CONTROLO	57
TABELA 6.8: TABELAS DE ORIGEM DOS DADOS PARA AS DIMENSÕES DO MÓDULO QIO	59
TABELA 6.9: TABELAS DE ORIGEM DOS DADOS PARA OS INDICADORES DO MÓDULO QIO	60
TABELA 6.10: ESTIMATIVA EMPÍRICA DO ESPAÇO A OCUPAR PELA BASE DE DADOS DE ANÁLISE E CONTROLO POR ANO	63
TABELA 6.11: MATRIZ DE CARACTERIZAÇÃO DAS FERRAMENTAS OLAP (TRADUZIDO DE [PENDSE & CREETH, 2000A]).....	65
TABELA 6.12: AVALIAÇÃO DAS FERRAMENTAS <i>OLAP</i>	66
TABELA 6.13: ESPAÇO OCUPADO POR UM ANO DE DADOS (FOLHA DE CÁLCULO EXPLICADA EM [MICROSOFT CORPORATION, 1998]).....	90

1 Introdução

O trabalho apresentado teve como principal objectivo a construção de um sistema de Armazenamento de Dados com base num método de desenvolvimento adequado. Foi utilizado como caso de estudo o sistema GIST98/EUROBUS - um Sistema de Apoio à Decisão para o planeamento de empresas de transportes colectivos rodoviários de passageiros.

A importância dos sistemas de Armazenamento de Dados é hoje incontestável, reflectindo-se essa importância numa evolução muito acentuada, quer na diversidade de produtos oferecida quer na inovação técnica, quer ainda no volume de negócios associados, como o demonstra a Figura 1.1.

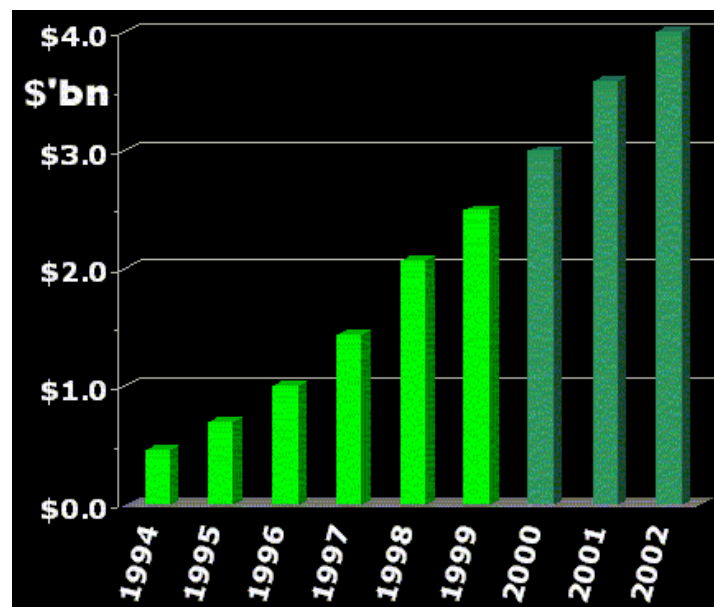


Figura 1.1: O mercado mundial de ferramentas OLAP em milhares de milhões de US\$ (JPendse & Creeth, 2000a)

Pode-se mesmo dizer que os sistemas de Armazenamento de Dados estão na moda, moda essa justificada pelo reconhecimento por parte do mercado das mais valias que se podem obter com a transformação dos dados em informação.

Porque é também e sobretudo disso que os sistemas de Armazenamento de Dados tratam: a transformação dos dados em conhecimento útil para o negócio transformando-os numa vantagem competitiva.

1.1 Organização da dissertação

A dissertação está dividida em duas grandes partes: na primeira parte, constituída pelos capítulos 2 e 3, são definidos os conceitos fundamentais e apresentados métodos para o desenvolvimento de sistemas de Armazenamento de Dados; na segunda parte é feita, no capítulo 4, a apresentação do caso de estudo, no capítulo seguinte é descrito o método de desenvolvimento utilizado, e no 6º capítulo descreve-se a aplicação desse método ao caso de estudo. Pretende-se com esta organização fazer uma separação clara entre o trabalho metodológico e o trabalho aplicacional tentando mostrar, com este último, a validade do primeiro.

Nas conclusões são abordadas algumas questões de carácter mais estratégico quer na perspectiva do sistema GIST98/EUROBUS quer na perspectiva de um sistema de Armazenamento de Dados.

A documentação essencial para um melhor acompanhamento da leitura desta dissertação surge em anexos. Assim, no anexo A encontram-se as propostas de indicadores efectuadas pelas empresas, e no anexo B mostram-se os esquemas da Bases de Dados de planeamento feitos com recurso à ferramenta *ORACLE Designer 2000*.

1.2 Algumas considerações sobre a elaboração da dissertação

Dada a complexidade técnica desta área de trabalho, é difícil tratar com a profundidade desejável os vários temas que estão subjacentes a um trabalho desta natureza. Assim, muitas das questões processuais e técnicas, que se colocaram, foram, por diversas vezes, abordadas numa perspectiva essencialmente pragmática em detrimento de uma análise exaustiva. Ainda devido à complexidade deste tema, assumiu-se a familiaridade do leitor com a maioria dos conceitos associados a Bases de Dados relacionais [Date, 1990].

Os dados apresentados foram propositadamente alterados para garantir a sua confidencialidade.

Na escrita desta dissertação prestou-se especial atenção a um aspecto que é da maior relevância neste tipo de trabalhos, particularmente para quem desenvolve actividades de ensino: a preservação da língua portuguesa através da pesquisa dos termos equivalentes aos utilizados em língua inglesa (trabalho muitas vezes penoso). No entanto, com o objectivo de manter a coerência no desenvolvimento dos vários módulos do sistema

GIST98/EUROBUS, utilizou-se a língua inglesa no código desenvolvido, nomeadamente nos nomes de tabelas, variáveis ou outros.

Uma última palavra para as notações adoptadas na escrita:

- Palavras de língua inglesa que já entraram na nossa forma corrente de expressão, assim como nomes de empresas, produtos, variáveis ou termos técnicos, aparecem em itálico, como por exemplo *software*, *NCR Corporation*, *Knosys ProClarity 2.0*, *import_date* ou *triggers*, respectivamente.
- Conceitos relevantes na área em estudo aparecem com a primeira letra de cada palavra em maiúscula, como por exemplo Armazém de Dados. O mesmo acontece para as palavras Tabela e Figura.
- Palavras ou expressões em que uma parte tem um valor fixo e outra parte um valor variável, aparecem da seguinte forma *valor_fixo<valor_variável>*, como por exemplo *map_<nome da entidade>* em que uma possível concretização é *map_paths*.

2 Conceitos fundamentais

Neste capítulo definem-se os conceitos fundamentais associados ao Armazenamento de Dados e faz-se uma caracterização dos tipos de arquitectura existentes no mercado.

2.1 Definições

2.1.1 Sistema de Apoio à Decisão

“Um Sistema de Apoio à Decisão é um sistema de informação em suporte informático, interactivo, flexível e adaptável, especialmente desenvolvido para apoiar a resolução de problemas de gestão não estruturados, melhorando o processo de tomada de decisão associado. Um tal sistema utiliza dados, fornece uma interface fácil com o utilizador e pode adaptar-se às opções de quem decide. Um Sistema de Apoio à Decisão pode ainda utilizar modelos, é construído por um procedimento interactivo (muitas vezes pelo utilizador final), apoia todas as fases de tomada de decisão e pode incluir ainda componentes de conhecimento”.

Esta definição [Turban & Aronson, 1998], traduzida de forma livre, tem, na perspectiva de quem trabalha com dados, duas componentes distintas: por um lado, os Sistemas de Apoio à Decisão requerem dados, por outro, os dados e a informação neles contida servem para apoiar os agentes de decisão na resolução de problemas não estruturados por forma a melhorar a tomada de decisão. É determinante compreender de forma clara a finalidade dos dados. Se a finalidade fosse outra, diferente da que atrás foi referida, por exemplo a constituição de arquivos documentais, a problemática do que vai ser tratado de ora em diante seria completamente diversa.

2.1.2 Armazém de Dados (*Data Warehouse*)

Para [Kelly, 1996],

“O Armazém de Dados de uma empresa é o repositório de dados único e integrado, formando a infra-estrutura base para aplicações de *software* de informação da empresa.”,

ou seja, o Armazém de Dados é o portão que separa o mundo do processamento operacional do mundo dos Sistemas de Apoio à Decisão.

[Inmon, 1996] define Armazém de Dados como sendo:

A construção de um sistema de Armazenamento de Dados no âmbito do sistema GIST98/EUROBUS

“Um conjunto de dados temáticos, integrados, não voláteis e com o histórico da evolução ao longo do tempo, para o apoio às tomadas de decisão de gestão.”

É importante detalhar estas características:

- Temáticos, ou organizados por temas, pressupõe que a organização dos dados é dependente da forma como estes vão ser utilizados;
- Integrados, significa que existem definições únicas e consistentes para os dados, ou seja, o Armazém de Dados deve integrar toda a informação relevante para os fins a que se destina, garantindo a sua consistência e unicidade;
- Não voláteis, significa que uma vez introduzidos no Armazém de Dados, os dados não podem ser alterados ou eliminados;
- Histórico da evolução ao longo do tempo, pressupõe a possibilidade de obter a mesma informação para diferentes momentos temporais, ou seja, permite a análise da evolução da informação ao longo do tempo.

Ora, estas características não se encontram nas bases de dados operacionais, o que leva à necessidade de criar um repositório de dados próprio.

Dados operacionais vs. dados de informação

Inmon apresenta de forma detalhada as diferenças entre os dados operacionais e os dados de informação, das quais se realçam as apresentadas na Tabela 2.1.

Dados operacionais	Dados de informação
Orientados para as aplicações	Orientados para os utilizadores
Detalhados	Agregados
Exactos, referentes ao momento de acesso	Representam valores ao longo do tempo
Servem pessoas com cargos de responsabilidade mais baixa	Servem pessoas com cargos de responsabilidade mais alta
Podem ser alterados	Não podem ser alterados
Acede a um registo de cada vez	Acede a um conjunto de registos de cada vez
Orientados para a transacção	Orientados para a análise
Não redundantes	Redundantes
Estrutura estática	Estrutura flexível
Pequena quantidade de dados utilizada num processo	Grande quantidade de dados utilizada num processo

Tabela 2.1: Comparação entre os dados operacionais e os dados de informação (traduzido de [Inmon, 1996])

2.1.3 Armazenamento de Dados (*Data Warehousing*)

Pelas definições acima apresentadas, o Armazém de Dados compreende apenas os dados, não havendo qualquer referência a todos os processos envolvidos na sua gestão. Assim, passou a utilizar-se o termo Armazenamento de Dados para focar a definição nos processos e não nos dados.

Utilizando a definição dada por [Watterson, 1997]:

“Armazenamento de Dados refere-se a um conjunto de processos ou a uma arquitectura que junta dados relacionados oriundos de diferentes sistemas operacionais de forma a dar uma visão integrada dos dados que pode atravessar vários departamentos da empresa.”

2.1.4 Armazém de Dados Departamental (*Data Mart*)

Existem duas concepções diferentes sobre o que é um Armazém de Dados Departamental:

- Na primeira visão, o Armazém de Dados Departamental é um subconjunto do Armazém de Dados com a informação relevante para um departamento ou uma função, estabelecendo dessa forma uma descontinuidade natural nos dados. Esta visão pressupõe que o Armazém de Dados já existe e que, portanto, o Armazém de Dados Departamental não é mais do que um subconjunto extraído de um todo integrado: o Armazém de Dados.
- A segunda visão surgiu da dificuldade sentida por muitas empresas em desenhar e desenvolver de raiz um Armazém de Dados, uma vez que a complexidade e o custo envolvidos são significativos, e que tais projectos se constituem tipicamente como projectos de média ou longa duração. Assim, o Armazém de Dados Departamental é desenvolvido tendo em consideração as necessidades de um departamento ou de uma determinada função, sendo os vários Armazéns de Dados Departamentais desenvolvidos individualmente e desfasados no tempo. Neste caso, o Armazém de Dados será a reunião dos vários Armazéns de Dados Departamentais não garantindo que os dados sejam integrados, característica determinante na definição de Armazém de Dados (subsecção 2.1.2).

No primeiro caso diz-se que o Armazém de Dados Departamental é dependente, enquanto no segundo caso se diz independente.

Armazém de Dados vs. Armazém de Dados Departamental

Do que atrás se disse, pode-se facilmente depreender que um Armazém de Dados Departamental é diferente de um Armazém de Dados. Para melhor perceber o alcance dessas diferenças observe-se a Tabela 2.2.

	Armazém de Dados Departamental	Armazém de Dados
Tamanho¹	Média: 50 GB; Máx: ~250GB	Pode atingir vários TB
Propósito	Específico a uma área/função	Repositório da empresa
Abordagem	De baixo para cima	De cima para baixo de acordo com o modelo de dados da empresa
Controlo	Departamento	Sistemas de informação
Prazo	3 a 6 meses	1 a 2 anos

Tabela 2.2: Comparação entre os Armazéns de Dados Departamentais e os Armazéns de Dados (traduzido de [Watterson, 1997])

2.1.5 Granulosidade

“Granulosidade refere-se ao grau de detalhe com que os dados são guardados no Armazém de Dados. Quanto maior é o detalhe, menor é a granulosidade. Quanto menor é o detalhe, maior é a granulosidade.” [Inmon, 1996]

Tal como se viu na tabela Tabela 2.1, os dados de informação são agregados, enquanto os dados operacionais são detalhados. Assim, num Armazém de Dados a granulosidade dos dados é uma característica fundamental a definir, dado que afecta em simultâneo a dimensão do Armazém de Dados e o potencial de informação que lhe está associado. Quanto menor a granulosidade dos dados, maior será o Armazém de Dados (aumentando a complexidade técnica e, consequentemente, os custos de implementação e manutenção) permitindo, no entanto, análises mais detalhadas por parte dos agentes de decisão.

2.1.6 Multi-dimensionalidade: conceitos associados

Indicador

Indicador é um atributo variável ao longo do tempo, normalmente numérico, que é relevante como instrumento de análise para a empresa.

¹ Estes valores são meramente indicativos, podendo existir Armazéns de Dados Departamentais com tamanho superior a outros Armazéns de Dados. É, no entanto, pouco razoável admitir que isso possa acontecer dentro da mesma empresa.

Dimensão

Designa-se por dimensão uma estrutura hierárquica agrupada por níveis, sendo cada um desses níveis constituído por membros. Os membros são, normalmente, dados alfanuméricos. Por exemplo, a dimensão tempo constituída pelos níveis ano, trimestre, mês e dia, onde o ano é definido com sendo um conjunto de trimestres, que por sua vez é um conjunto de meses constituídos por um conjunto de dias. O ano de 1999 é um membro do nível ano, enquanto 2 de Junho de 1999 é um membro do nível dia, que está associado ao membro Junho de 1999 do nível mês, que por sua vez está associado ao membro 2º trimestre de 1999 do nível trimestre que, por último, está associado ao membro 1999 do nível ano. Note-se que não é possível definir semana como um nível da dimensão tempo hierarquicamente inferior ao nível mês, dado que uma semana pode pertencer a um mês, não por inteiro, mas apenas parcialmente.

Cubo (também designado por hiper-cubo ou estrutura multi-dimensional)

Um Cubo é definido em [OLAP Council, 1995] como sendo “um grupo de células com dados dispostos em função das suas dimensões”. Apesar de um cubo, tal como é vulgarmente designado, ser uma figura geométrica tridimensional com arestas de igual tamanho, o Cubo, como estrutura multi-dimensional, tem um significado mais amplo, podendo ter um número variável de dimensões de diferentes tamanhos. Por exemplo, uma folha de cálculo exemplifica um Cubo de duas dimensões com as células definidas por linhas e colunas, onde tanto os cabeçalhos das linhas como das colunas são membros de cada uma dessas dimensões.

Agregação (também designada por consolidação)

Consiste no processamento dos dados de forma a obter valores referentes a níveis hierárquicos superiores. Por exemplo, se existir informação da facturação diária, é necessário efectuar a agregação (ou consolidação) dos dados usando, neste caso, a operação de agregação soma para obter a facturação mensal, trimestral ou anual.

Dados detalhados

São os dados associados aos membros dos níveis inferiores das dimensões. Usando o exemplo anterior, apenas os dados referentes à facturação diária são detalhados.

Dados agregados (também designados por dados consolidados)

Por oposição à definição de dados detalhados, são os dados associados aos restantes membros das dimensões. São obtidos a partir de dados detalhados por agregação. Usando o exemplo anterior, os dados das facturações mensais, trimestrais e anuais são agregados, já que são obtidos a partir de dados detalhados utilizando uma função de agregação (neste caso, a função soma).

2.1.7 OLAP - processamento analítico por controlo directo (*On-Line Analytical Processing*)

O conceito foi definido em 1993 por E. F. Codd num artigo patrocinado pela *Arbor Software* intitulado “Providing OLAP to user analysts: an IT mandate”. O conceito esteve sempre associado a contextos mais comerciais do que técnicos.

Como o nome indica, processamento analítico está relacionado com o processo de análise. De facto, o processamento analítico corresponde a efectuar operações de leitura sobre as Bases de Dados com o objectivo de obter dados para análise. Enquanto *OLTP* – processamento transaccional por controlo directo (*On-Line Transactional Processing*) conduziu a uma determinada organização dos dados (tipicamente Bases de Dados relacionais na 3ª forma normal² [Date, 1990]) como forma de otimizar as operações de transacção (inserção, actualização e remoção), a optimização das Bases de Dados para operações de leitura conduziu a uma mudança de paradigma na estruturação das Bases de Dados (ver subsecção 2.2.3).

Assim, usando a definição dada por [OLAP Council, 1995]:

“*OLAP* é um conjunto de ferramentas de *software* que permite aos analistas, gestores e executivos obter dados, através de um acesso rápido, consistente e interactivo a uma vasta variedade de possíveis vistas de informação, que foram transformadas a partir de dados fonte, de forma a reflectir a real dimensionalidade da empresa tal como é entendida pelo utilizador.

A funcionalidade *OLAP* é caracterizada pela análise multi-dimensional dinâmica de dados consolidados da empresa para apoio a actividades de “navegação” e análise por parte do utilizador final, incluindo:

² Existem regras no desenho de Bases de Dados relacionais que conduzem à minimização do espaço ocupado. Essas regras são designadas por formas normais. Quanto mais normalizada está a base de dados menos espaço ocupa, mas mais lentas são as operações de leitura.

- Cálculos e modelação aplicada, cruzando dimensões através de hierarquias e/ou atravessando membros;
- Análises de tendências através de períodos de tempo sequenciais;
- Selecção de subconjuntos para visualização em monitor;
- Detalhe da informação em níveis de consolidação mais baixos;
- ‘Navegação’ pelos dados detalhados;
- Rotação para comparações de novas dimensões na área de visualização.

O *OLAP* é implementado no modo cliente/servidor com multi-utilização, oferecendo respostas rápidas e consistentes às consultas, independentemente do tamanho e complexidade das Bases de Dados. O *OLAP* ajuda o utilizador a sintetizar a informação da empresa através da visualização comparativa e personalizada, assim como através da análise de dados históricos e de dados previstos em vários cenários de modelos de dados de previsão. Isto é alcançado através da utilização de um Servidor *OLAP*. ”

É importante notar que os termos Armazenamento de Dados e *OLAP* se sobrepõem parcialmente, facto que não é estranho a terem surgido em contextos diferentes. No entanto, a distinção dos conceitos pode ser entendida como uma diferença de perspectiva: enquanto no primeiro caso a perspectiva é colocada do lado dos processos necessários ao Armazenamento de Dados, no segundo caso a perspectiva é colocada nos processos necessários à análise dos dados [Thomsen, 1997]. Ou seja, no primeiro caso a perspectiva é posta nos processos necessários aos dados, enquanto no segundo caso a perspectiva está do lado dos processos necessários aos utilizadores.

2.1.8 Servidor *OLAP*

Utilizando, ainda, a definição dada por [OLAP Council, 1995]:

“Um Servidor *OLAP* é um sistema de manipulação de dados de alta capacidade e de multi-utilização, especificamente desenhado para suportar e operar com estruturas de dados multi-dimensionais. Uma estrutura multi-dimensional é formada de modo a que todos os dados sejam localizados e acedidos com base na intersecção dos membros das dimensões que definem esse dado específico. O desenho do servidor e a estrutura dos dados são optimizados para obtenção rápida de qualquer informação específica, assim como para cálculos rápidos e flexíveis e, ainda, para a transformação de dados fonte

com base em fórmulas de cálculo. O Servidor *OLAP* pode, também, tratar fisicamente a informação multi-dimensional processada, de forma a dar tempos de resposta rápidos e consistentes aos utilizadores finais, ou pode carregar as estruturas de dados em tempo real a partir de Bases de Dados relacionais ou outras, ou oferecer uma selecção de ambas. Dado o estado tecnológico actual e a exigência, da parte dos utilizadores finais, de tempos de resposta rápidos e consistentes, a opção de colocar os dados multi-dimensionais no Servidor *OLAP* é a mais utilizada.”

2.2 Arquitectura do Armazenamento de Dados

A arquitectura do Armazenamento de Dados [The Data Warehousing Institute, 1998] pode ser vista como constituída por três partes distintas (Figura 2.1): a aquisição dos dados, a administração das estruturas de informação e as ferramentas para o apoio à decisão. Esta última parte é composta por três grandes famílias de produtos: as soluções de análise multi-dimensional; as aplicações de utilização amigável e produção de relatórios; e as aplicações verticais e de mineração de dados. Um componente importante de um Armazém de Dados são os Meta Dados que têm um papel relevante na documentação e na recuperação de falhas.

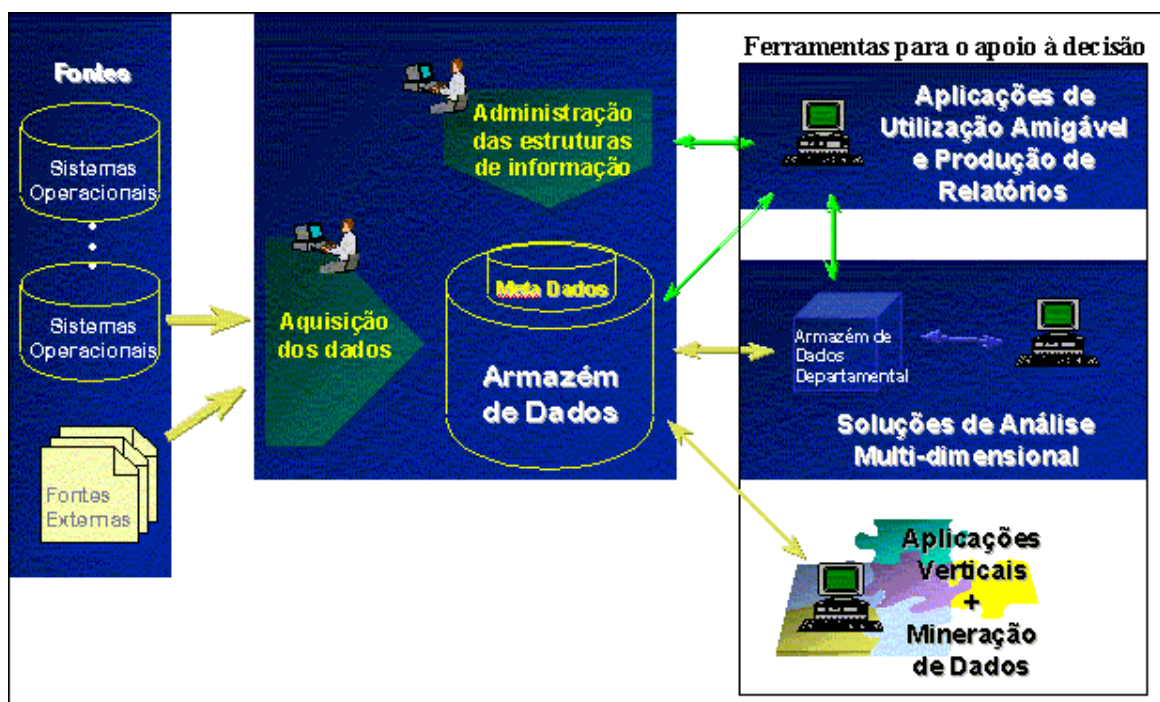


Figura 2.1: Arquitectura de referência do Armazenamento de Dados (adaptado de [Ranito, 1999])

2.2.1 Aquisição dos dados

O processo de aquisição dos dados compreende duas fases distintas: a purificação e a migração dos dados.

A fase de purificação dos dados, que garante a qualidade dos dados fonte, é composta por duas tarefas essenciais:

- O estudo da natureza e qualidade dos dados fonte, e
- O tratamento dos dados fonte de forma a torná-los consistentes.

A migração dos dados compreende as seguintes tarefas:

- Extração, que pode corresponder a operações de leitura simples ou envolver processos mais complexos de forma a seleccionar o conjunto de dados pretendido;
- Transformação, que permite aplicar regras de forma a integrar dados oriundos de diferentes tabelas e sistemas;
- Transporte, que é a passagem dos dados das fontes para o Armazém de Dados, sendo habitualmente feito por transferência de ficheiros;
- Carregamento, normalmente accionado pelas ferramentas de transformação de dados após a sua conclusão; de realçar o facto de as ferramentas de transformação dos dados permitirem habitualmente a sua formatação de acordo com a forma como estão definidos no Armazém de Dados.

O processo descrito tem por objectivo carregar o Armazém de Dados com dados que garantam certas características de qualidade tal como se pode ver na Tabela 2.3.

Características	Descrição
Preciso	Concordância entre os valores armazenados e os valores reais
Completo	Armazenamento de todos os dados necessários
Consistente	Concordância ou coerência lógica entre os dados libertando-os de desvios ou contradições
Relacionado	Concordância ou coerência lógica que permite uma correlação racional na comparação com outros dados similares
Atempado	Dados armazenados dentro do período de tempo especificado
Único	Adequação dos dados às restrições de unicidade existentes
Válido	Adequação dos dados aos tipos definidos no Armazém de Dados

Tabela 2.3: Características de qualidade (traduzido de [Hufford, 1996])

Convém, ainda, salientar o facto de o tempo de desenvolvimento do processo de aquisição dos dados ser percentualmente bastante elevado³ em relação ao tempo total de implementação de um sistema de Armazenamento de Dados.

2.2.2 Administração das estruturas de informação

A administração das estruturas de informação é feita recorrendo aos Sistemas de Gestão de Bases de Dados. De facto, estes evoluíram bastante nos últimos anos, em parte como resposta à necessidade de gerir grandes quantidades de informação (Bases de Dados com vários *Tera Bytes*) permitindo baixos tempos de resposta a operações de leitura. Assim, alguns dos Sistemas de Gestão de Bases de Dados utilizados para a gestão dos Armazéns de Dados são os antigos sistemas aos quais foram acrescentadas funcionalidades que lhes permitem gerir eficazmente grandes quantidades de informação.

Algumas dessas funcionalidades são:

- As partições - são tabelas divididas em partes, de acordo com um certo critério. Por exemplo, a tabela com as facturas pode ser dividida em várias partes, de acordo com o ano de emissão. Podem ser vistas como tabelas parametrizadas que, em função de um determinado parâmetro (no exemplo dado, o ano), guardam os dados na partição correspondente. Quando aparece o primeiro registo referente a um dado ano, é criada automaticamente a partição correspondente.
- Técnicas de indexação - surgiram várias técnicas de indexação novas que permitem operações de selecção mais rápidas, tais como os índices *bitmap* e outros. Os índices *bitmap* são particularmente úteis quando o número de registos é elevado, mas a gama de valores possíveis para um determinado campo é pequeno.
- Processamento paralelo - capacidade de tirar partido da existência de vários processadores, dividindo as consultas em sub-tarefas, de modo a reduzir os tempos de resposta.

Meta Dados

Os Meta Dados são dados acerca dos dados. Existem algumas soluções em que os Meta Dados são guardados num repositório de dados próprio, permitindo aos administradores do Armazém de Dados a sua utilização para tarefas de criação e manutenção, e dando a

³ Segundo profissionais com grande experiência nesta área, esse valor deve rondar os 80%. No entanto não foi possível encontrar estudos que o fundamentem.

possibilidade aos utilizadores finais de os utilizar como um catálogo do Armazém de Dados onde podem ver, entre outras coisas, as fórmulas utilizadas para o cálculo de um determinado indicador, a descrição de uma dimensão, etc. No entanto, é ainda muito comum encontrar soluções em que os Meta Dados são gerados automaticamente e guardados em formatos próprios, inacessíveis tanto aos administradores do Armazém de Dados como aos utilizadores finais servindo, neste caso, para guardar dados utilizados exclusivamente pelo Sistema de Gestão de Bases de Dados.

2.2.3 Ferramentas para o apoio à decisão

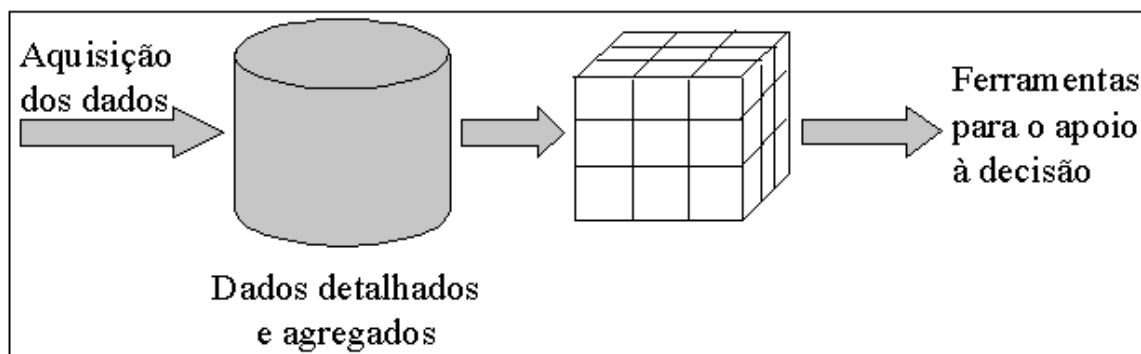
Análise multi-dimensional

A análise multi-dimensional requer a utilização de Servidores *OLAP*. Alguns Servidores *OLAP* suportam uma grande diversidade de tecnologias, outros são específicos para um determinado Sistema de Gestão de Bases de Dados.

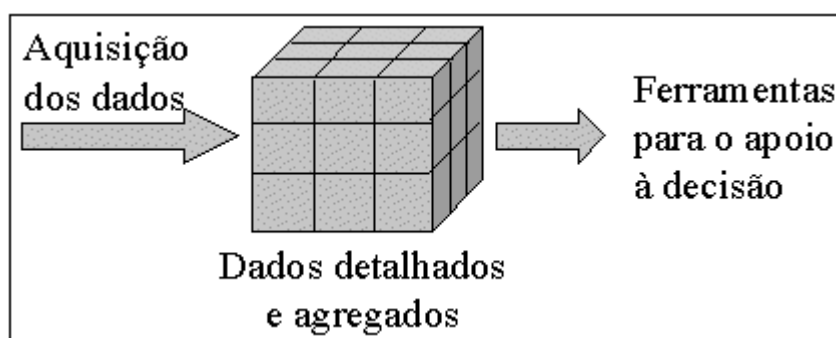
Existe um grande número de variáveis para a caracterização das arquitecturas utilizadas pelos diferentes Servidores *OLAP* existentes no mercado. De facto, existem tantas arquitecturas quantos os produtos. A caracterização das arquitecturas *OLAP* será feita utilizando dois critérios [Pendse & Creeth, 2000a]: o local onde ocorre a maioria do processamento multi-dimensional e a forma de armazenar os dados.

Quanto às formas de armazenamento existem quatro opções:

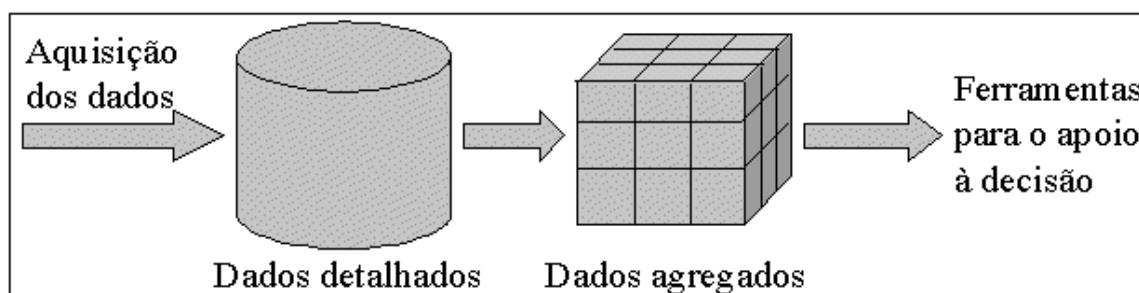
- Base de Dados relacional, que é uma opção comum, até porque os dados fonte são normalmente retirados de Bases de Dados relacionais. No entanto, para acelerar os tempos de resposta, a Base de Dados não é normalizada na 3ª forma normal. De facto, quanto mais normalizada está uma Base de Dados menos espaço ela ocupa, mas maiores são os tempos de resposta às operações de consulta. Existem esquemas específicos de desenho para diminuir os tempos de resposta das consultas nomeadamente os esquemas em estrela ou em cristal de gelo dos quais se falará com mais profundidade na subsecção 6.5.1. Os Servidores *OLAP* que utilizam esta forma de armazenamento usam tabelas temporárias para guardar os dados agregados. As arquitecturas *OLAP* cujas Bases de Dados são relacionais designam-se por *ROLAP* (Figura 2.2).

Figura 2.2: Arquitectura *ROLAP*

- Base de Dados multi-dimensional, na qual os dados são guardados na máquina servidor numa Base de Dados multi-dimensional, que se distingue de uma Base de Dados relacional pelo facto de armazenar os dados em vectores multi-dimensionais em vez de tabelas. Os dados tanto podem ser guardados em disco como em memória RAM. É habitual os dados terem que ser pré-processados. As arquitecturas *OLAP* com esta forma de armazenamento designam-se por *MOLAP* (Figura 2.3).

Figura 2.3: Arquitectura *MOLAP*

- Base de Dados híbrida que, tal como o nome indica, é uma solução intermédia entre as arquitecturas *ROLAP* e *MOLAP*. Os dados são guardados numa Base de Dados relacional, sendo as agregações calculadas e disponibilizadas numa Base de Dados multi-dimensional. Esta abordagem, designada por *HOLAP* (Figura 2.4), tem vindo a ganhar importância.

Figura 2.4: Arquitectura *HOLAP*

- Ficheiros localizados no cliente, solução que corresponde a parte dos dados serem colocados nas máquinas cliente, podendo ficar tanto em disco como em memória RAM. A passagem dos dados para a máquina cliente pode ser feita previamente ou a pedido. Designa-se esta opção por *DOLAP - OLAP* cliente (do inglês *Desktop OLAP*).

Quanto ao local onde ocorre a maioria do processamento multi-dimensional existem três opções:

- *SQL (Structured Query Language)* em passos múltiplos, na qual as operações de consulta são efectuadas através de vários comandos *SQL*, sendo os resultados guardados temporariamente na máquina cliente ou noutro local auxiliar e posteriormente aí processados, podendo ainda ser colocados em memória *cache*, o que aumenta significativamente a velocidade de resposta;
- no Servidor *OLAP*, a opção mais utilizada, sendo preponderante a existência de memória RAM em quantidade para a obtenção de bons tempos de resposta;
- no cliente *OLAP*, assumindo que as máquinas cliente têm boa capacidade de processamento, parte do cálculo multi-dimensional pode ser para elas transferido.

Aplicações de utilização amigável e produção de relatórios

São aplicações a que o utilizador final tem acesso. Permitem a análise dos dados e a produção de relatórios, sendo de fácil utilização.

Aplicações verticais e de mineração de dados

Estas aplicações permitem a descoberta semi-automática de padrões, alterações, associações e anomalias em grandes quantidades de dados. Utilizam métodos estatísticos, modelos e uma grande diversidade de algoritmos.

3 Métodos de desenvolvimento

Para se poder analisar a adequação de um determinado método a um dado problema é necessário, antes de mais, perceber quais os factores críticos de sucesso na resolução desse problema. A boa compreensão das principais causas de insucesso é particularmente relevante, salientando [Ebel, 1998] que cerca de metade dos projectos de Armazenamento de Dados falham.

Assim, começa-se por apresentar os principais factores críticos para o sucesso de um projecto de Armazenamento de Dados; de seguida, descrevem-se três métodos de desenvolvimento e, por último, é feita uma breve análise final.

3.1 Factores críticos de sucesso

Como é evidente, não existe uma lista única de factores críticos de sucesso. De facto, a cada experiência de desenvolvimento de um projecto de Armazenamento de Dados estará associado um conjunto de factores críticos para o sucesso desse projecto que variará de projecto para projecto e será com certeza diferente de autor para autor. A lista aqui apresentada tem como único objectivo alertar para alguns dos factores mais críticos que, quando não se verificam, podem conduzir ao fracasso de um projecto de Armazenamento de Dados. Tal lista não pretende ser exaustiva.

Os factores críticos de sucesso podem ser separados em dois grupos: os que estão relacionados com a condução (ou gestão) dos projectos e aqueles que dizem respeito à competência técnica para dar bom seguimento a esses projectos.

Do primeiro grupo destacam-se [Paul, 1997], [Perkins, 1998], [Onder & Nash, 1998]:

- Apoio de alto nível – O envolvimento e apoio de gestores com poder de decisão ao mais alto nível são fundamentais. Projectos desta natureza obrigam a uma grande disponibilidade por parte de muita gente da empresa cliente, assim como o acesso a muita informação. Sem o apoio adequado de alguém com influência dentro da empresa, o sucesso fica seriamente comprometido. O projecto deve ser encarado como uma forte aposta estratégica por parte da empresa.
- Gestão das expectativas – Particularmente nos projectos de média/longa duração tem que haver um compromisso entre o trabalho de desenvolvimento e a divulgação desse

mesmo trabalho, sob risco de defraudar as expectativas geradas pelo cliente. A objectividade e o pragmatismo devem ser uma constante. É preferível obter resultados menos ambiciosos, encurtando o tempo de apresentação ao cliente, do que tentar soluções mais complexas, caso isso possa atrasar a referida apresentação. A elaboração de planos de trabalho e especificações de desenvolvimento permitem, entre outras coisas, objectivar o trabalho e torná-lo mais perceptível para o utilizador.

- Desenvolvimento orientado para o negócio – Sendo uma ferramenta para o apoio à tomada de decisão, é fundamental que a informação disponível seja pertinente para o negócio em causa. O desenvolvimento de um sistema de Armazenamento de Dados tem que partir, por parte da equipa de desenvolvimento, de um bom conhecimento do negócio.
- Constituição da equipa de desenvolvimento – Tal equipa tem que ser formada por um conjunto de especialistas com bons conhecimentos técnicos, capazes de entender a importância da informação para o negócio e com prática em projectos de Armazenamento de Dados. Empenho e dedicação não são suficientes neste tipo de projecto. O método de desenvolvimento utilizado é crítico. Não havendo um método bem delineado, por muita competência técnica que exista, o insucesso será o resultado mais provável.

Quanto aos factores críticos mais directamente relacionados com as competências técnicas da equipa de desenvolvimento salienta-se [Lee, 1998]:

- Qualidade dos dados - Garantir a qualidade dos dados é tecnicamente a tarefa mais árdua, sendo crítica a sua resolução. A perda de confiança na correcção dos dados por parte do utilizador conduz ao progressivo abandono do sistema e à procura de processos alternativos.
- Facilidade de utilização - Quanto mais intuitiva for a utilização das ferramentas e da informação disponibilizadas ao utilizador final maior será o número de utilizadores do sistema. A definição dos Cubos, em particular a associação das dimensões aos indicadores respectivos, deve ser especialmente cuidada para garantir que a informação esteja organizada de uma forma intuitiva para os utilizadores finais.

3.2 Apresentação dos métodos

Existem vários métodos de desenvolvimento [Watterson, 1997], [Adelman & Oates, 1998] para projectos de Armazenamento de Dados, quase sempre associados a empresas de consultoria especializadas nesta área de negócios. Nesta secção serão analisados três desses métodos de outras tantas empresas. Não é intenção deste capítulo fazer uma análise exaustiva dos métodos existentes e muito menos seleccionar o melhor desses métodos, até porque a informação disponível é, em alguns casos, insuficiente, o que se justifica pelo facto de serem propriedade intelectual das respectivas empresas e, assim, existirem reservas na sua divulgação. Pretende-se, apenas, perceber como se deve abordar um problema desta natureza.

3.2.1 Método Hadden – Kelly [Hadden & Company, 1999]

O método *Hadden-Kelly* assenta em cinco princípios:

- O Armazém de Dados deve estar alinhado com a estratégia da empresa cliente;
- A melhoria dos sistemas de análise de informação requer a combinação de conhecimentos técnicos e de negócio;
- Um relatório com a especificação inicial é essencial para assegurar a integração dos resultados e responder às necessidades de mudança do negócio;
- O Armazém de Dados e/ou Armazém de Dados Departamental deve ser construído de forma incremental, em que cada incremento traga um significativo benefício para o negócio, devendo esse benefício ser mensurável;
- O desenvolvimento deve ser rápido e focado.

O método *Hadden-Kelly* é apresentado como um processo iterativo composto por quatro etapas (Figura 3.1):

- Preparação: compreensão da organização face aos factores críticos de sucesso. Tem apenas uma fase:
 - Teste de aptidão: definição das condições de trabalho, nomeadamente através da obtenção do apoio das pessoas certas dentro da empresa cliente; e preparação de um plano para a etapa de planeamento.

Tem a duração de aproximadamente 2 semanas.

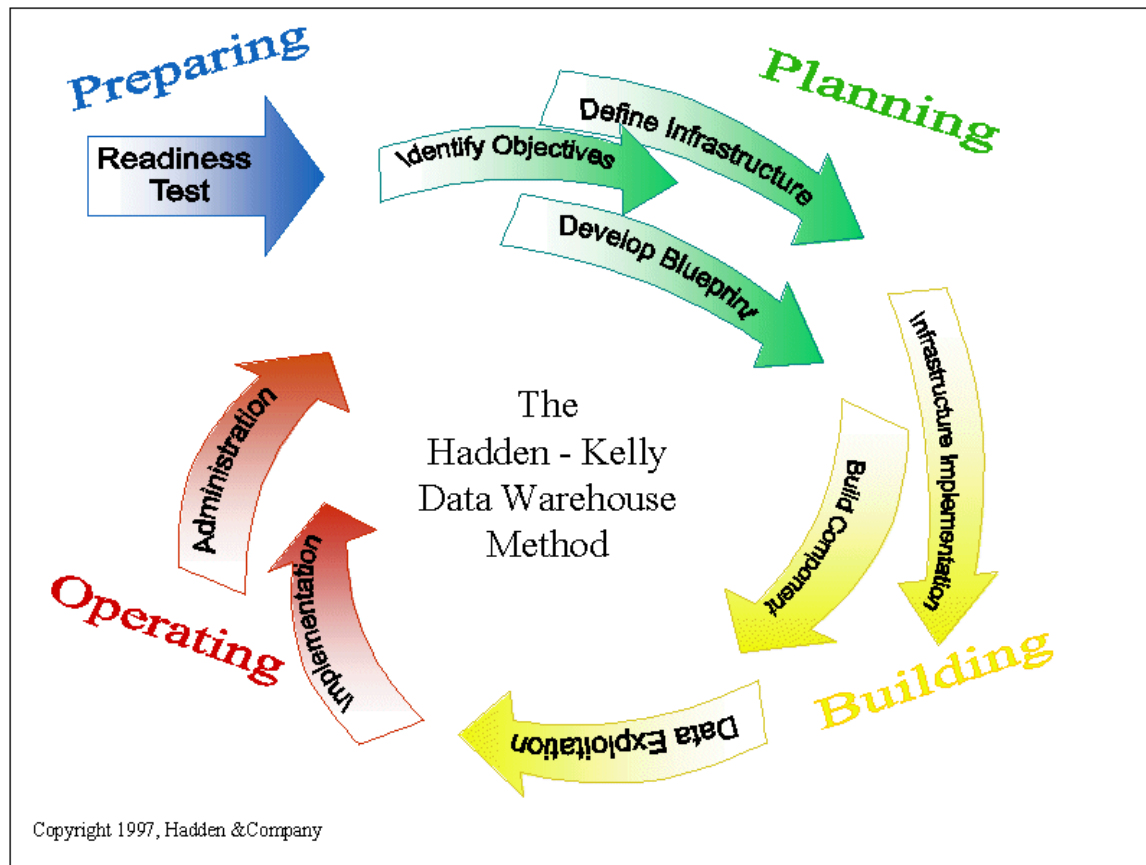


Figura 3.1: Método *Hadden-Kelly* ([Hadden & Company, 1999])

- Planeamento: identificação das necessidades de informação por parte da gestão para alcançar os objectivos de negócio e a definição da infra-estrutura requerida para a realização desses objectivos. Tem três fases:
 - ❑ Identificação dos objectivos de negócio: identificação e documentação das estratégias de negócio; organização das estratégias de acordo com o organograma da empresa; definição dos objectivos e dos indicadores e estabelecimento das prioridades.
 - ❑ Desenvolvimento do plano de implementação: identificação dos dados necessários para alcançar os objectivos; definição de regras de implementação; e elaboração do plano de implementação.
 - ❑ Definição da infra-estrutura de Armazenamento de Dados: identificação dos recursos técnicos e humanos necessários, assim como o desenvolvimento de regras e procedimentos de controlo da implementação.

Tem a duração de 2 a 3 meses.

- Construção: Desenvolvimento do Armazém de Dados e, eventualmente, de Armazéns de Dados Departamentais, desde que tal seja realizável no período de tempo definido para uma iteração. Tem três fases:
 - Implementação da infra-estrutura de Armazenamento de Dados: aquisição, instalação e desenvolvimento de competências com as ferramentas a utilizar na implementação;
 - Desenho e construção do Armazém de Dados: desenho do modelo de dados, desenvolvimento da aquisição dos dados e implementação da Base de Dados;
 - Análise e exploração dos dados: desenvolvimento das estruturas multi-dimensionais (definição dos cubos), instalação e teste das ferramentas necessárias de análise e exploração dos dados e produção de relatórios.

Tem a duração de 3 a 4 meses.

- Operação: Assegurar que o Armazém de Dados e/ou Armazéns de Dados Departamentais são eficazes face aos objectivos de negócio. É composto por duas fases:
 - Implementação do sistema de Armazenamento de Dados: instalação no cliente final, formação e avaliação da adequação do sistema aos objectivos de negócio;
 - Administração do sistema de Armazenamento de Dados: gestão dos Meta Dados, políticas de segurança, procedimentos de configuração, procedimentos de emergência, monitorização do desempenho, etc.

Esta etapa compreende tarefas de manutenção que são necessárias para o bom funcionamento do sistema. Assim, não existe prazo para a sua conclusão.

O método *Hadden-Kelly* dá especial ênfase à componente processual inerente a projectos desta natureza, nomeadamente através da produção de planos e relatórios como forma de preparação e síntese das várias etapas.

3.2.2 Método SDW [Ebel, 1998], [Gardner, 1998]

O método SDW da *NCR Corporation* é, também, iterativo mas, em vez das quatro etapas do método *Hadden-Kelly*, é composto por apenas três etapas tal como se pode ver na Figura 3.2, devendo cada iteração ter uma duração de 3 a 6 meses. A justificação para a opção por um método iterativo fica bem clara no paradoxo de Ebel:

“Se se tentar desenhar a versão final de um Armazém de Dados, provavelmente nem sequer se chega a iniciar a construção.”

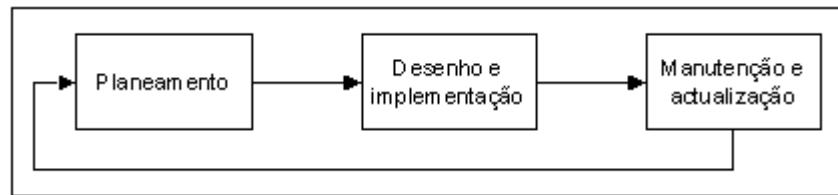


Figura 3.2: Método SDW (traduzido de [Ebel, 1998])

Antes de iniciar as etapas, é necessário efectuar algumas considerações prévias:

- As etapas que constituem cada uma das iterações devem estar alinhadas com o método seguido na gestão do projecto, tal como se pode ver na Figura 3.3.

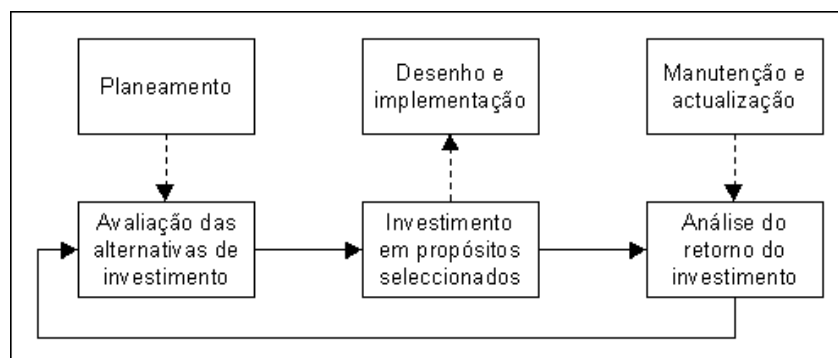


Figura 3.3: Método SDW de gestão (traduzido de [Ebel, 1998])

- O projecto não deve ser visto, por parte de quem o promove, como um produto, sendo da responsabilidade da equipa de desenvolvimento a capacidade de o perspectivar como uma forma de maximização do negócio, nomeadamente através da identificação de oportunidades para:
 - ❑ Redução de custos: optimização dos recursos, detecção de fraudes;
 - ❑ Novos negócios: novos produtos, novos mercados;
 - ❑ Fidelização dos clientes: programas de fidelização, prevenção na perda de clientes;
 - ❑ Gestão do negócio: *benchmarking*, controlo por objectivos;
 - ❑ Diagnósticos: correlação causa/efeito, origens de falhas.
- O âmbito do projecto a definir pela equipa de desenvolvimento deve ser simultaneamente ambicioso, em termos de objectivos, e cauteloso, na medida em que não deve exceder as limitações da equipa, sob risco de o projecto falhar.

- A equipa de trabalho deve incluir pessoas com experiência neste tipo de projecto na equipa de desenvolvimento assim como assegurar a participação efectiva por parte da empresa, nomeadamente através da inclusão das pessoas adequadas em cada fase do projecto.

Asseguradas estas considerações deve-se, então, passar às etapas propriamente ditas (Figura 3.4):

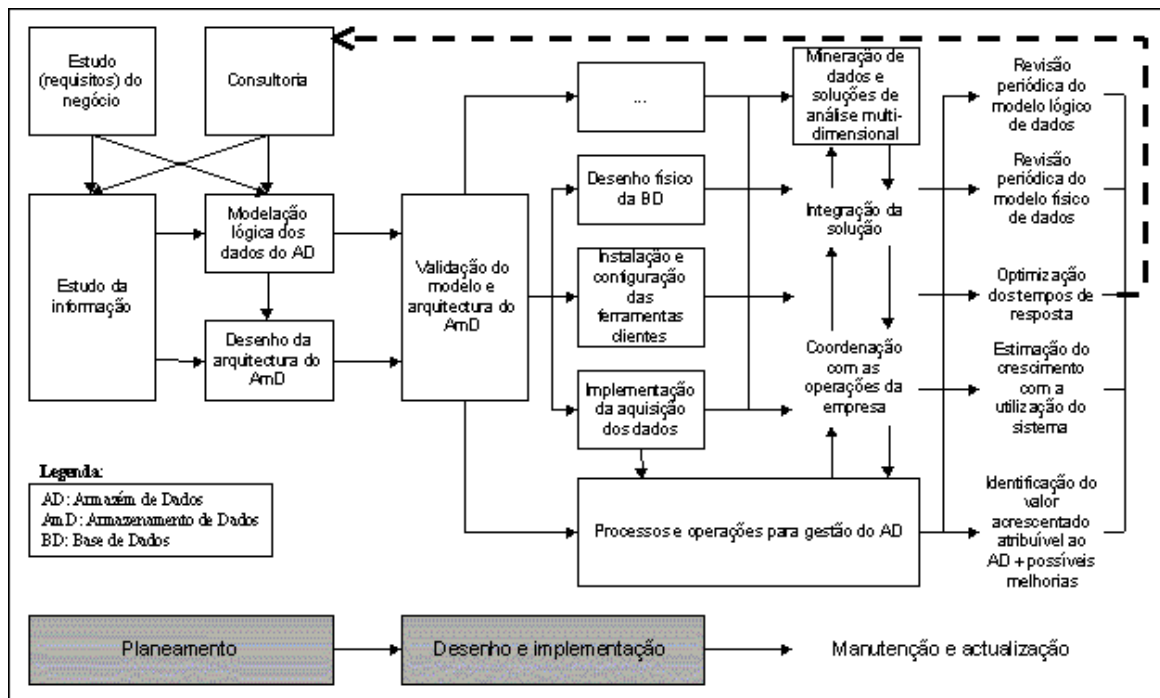


Figura 3.4: Método SDW detalhado (traduzido de [Gardner, 1998])

- **Planeamento:** Compreende todas as tarefas necessárias ao levantamento das necessidades de informação da empresa e o desenho do modelo lógico de dados e da arquitectura do Armazenamento de Dados. Desenvolve-se nas seguintes fases:
 - ❑ **Estudo do negócio:** identificação das áreas de negócio prioritárias, dos aspectos a melhorar e quantificação dos benefícios. Identificação das possíveis cooperações entre departamentos/funções.
 - ❑ **Consultoria:** papel essencialmente pedagógico com vista a sensibilizar a empresa para os benefícios de sistemas de Armazenamento de Dados.
 - ❑ **Estudo da informação:** Refinamento dos requisitos de negócio críticos e identificação do modo de obter a informação pretendida com os dados existentes.

- ❑ Modelação lógica dos dados do Armazém de Dados: elaboração do modelo lógico de dados, identificação das fontes, documentação de regras e definições.
- ❑ Desenho da arquitectura do Armazenamento de Dados: deve terminar com a elaboração de uma especificação que servirá de base para a etapa de desenho e implementação, devendo incluir a localização física do Armazém de Dados e Armazéns de Dados Departamentais, fontes e qualidade dos dados, aquisição dos dados e *hardware* e *software* necessários.
- Desenho e implementação: inclui o desenho físico do Armazém de Dados e a implementação de todos os processos associados ao Armazenamento de Dados. É constituída pelas seguintes fases:
 - ❑ Validação do modelo e arquitectura do Armazém de Dados: permite confrontar o modelo lógico de dados e a arquitectura de Armazenamento de Dados realizados com as realidades da equipa de desenvolvimento e da empresa cliente. No final desta fase o grau de confiança no modelo e na arquitectura planeada deve ser elevado.
 - ❑ Desenho físico da Base de Dados: o desenho de uma Base de Dados para um sistema de Armazenamento de Dados requer critérios distintos daqueles que se utilizam no desenho de Bases de Dados operacionais, nomeadamente no que diz respeito à normalização. Forte experiência no desenho de Bases de Dados relacionais na 3ª forma normal não é sinónimo de sucesso no desenho de Bases de Dados para sistemas de Armazenamento de Dados.
 - ❑ Instalação e configuração das ferramentas clientes: compreende a instalação e configuração das ferramentas para o apoio à decisão assim como o eventual desenvolvimento de aplicações necessárias ao utilizador final para o apoiar no processo de tomada de decisão.
 - ❑ Implementação da aquisição dos dados: pode ser feita recorrendo a ferramentas existentes ou ser desenvolvido de raiz.
 - ❑ Processos e operações para a gestão do Armazém de Dados: cópias de segurança, arquivo, replicação, recuperação de emergência, políticas de acesso, manutenção e actualização do *software* e *hardware*, administração dos utilizadores, monitorização

e processos necessários para a correcta manutenção do Armazém de Dados, etc. Esta fase prolonga-se na etapa seguinte.

- Manutenção e actualização: a última etapa é, ao contrário das anteriores, inteiramente realizada na empresa cliente compreendendo as seguintes fases:
 - ❑ Integração da solução: instalação de ferramentas e aplicações, configuração e testes finais. Deve ser executada em conjunto com técnicos da empresa cliente.
 - ❑ Coordenação com as operações da empresa: integração das operações de iniciação e finalização do *hardware* utilizado pelo sistema, assim como a sua manutenção e actualização.
 - ❑ Mineração de dados e soluções de análise multi-dimensional: instalação das ferramentas e aplicações necessárias, configuração e testes finais. Deve ser executada em conjunto com técnicos da empresa cliente.
 - ❑ Revisão periódica do modelo lógico de dados: esta fase é levada a cabo quando existem elementos da empresa cliente que têm experiência no desenho de Bases de Dados. É o esforço conjunto de encontrar o melhor compromisso na definição do modelo lógico de dados.
 - ❑ Revisão periódica do modelo físico de dados: idêntica à fase anterior, aplicando-se não ao modelo lógico de dados, mas sim ao modelo físico de dados.
 - ❑ Optimização dos tempos de resposta: a realizar por parte dos técnicos da empresa cliente (caso existam) quando o tempo de resposta não é satisfatório. Pode implicar alterações na configuração da rede, no sistema operativo, no Sistema de Gestão de Bases de Dados, etc.
 - ❑ Estimação do crescimento na utilização do sistema: projecção do crescimento na utilização do sistema e estudo e orçamento das alterações necessárias.
 - ❑ Identificação do valor acrescentado atribuível ao Armazém de Dados e de possíveis melhorias.

3.2.3 Método da DataWing Consulting Services, LLC [Welch, 1998]

“A construção de um Armazém de Dados é uma arquitectura e não uma tecnologia.”

Este lema, utilizado pela empresa de consultoria *DataWing Consulting Services, LLC*, pretende transmitir a ideia de que o processo de construção de um Armazém de Dados é uma arte e não um conjunto de técnicas quantificáveis que podem ser aplicadas através de uma sequência rígida de passos.

O método proposto também é iterativo sendo composto por seis etapas. A etapa inicial consiste na definição do âmbito do projecto, a análise consiste no levantamento daquilo que existe e daquilo que é necessário, o desenho corresponde à planificação do sistema de Armazenamento de Dados, sendo a etapa de desenvolvimento a construção desse mesmo sistema de acordo com o desenho realizado. A etapa de teste tem como objectivo validar o sistema construído, sendo a etapa de implementação aquela em que se põe o sistema a funcionar na empresa cliente. Estas fases são decompostas da forma apresentada na Figura 3.5.

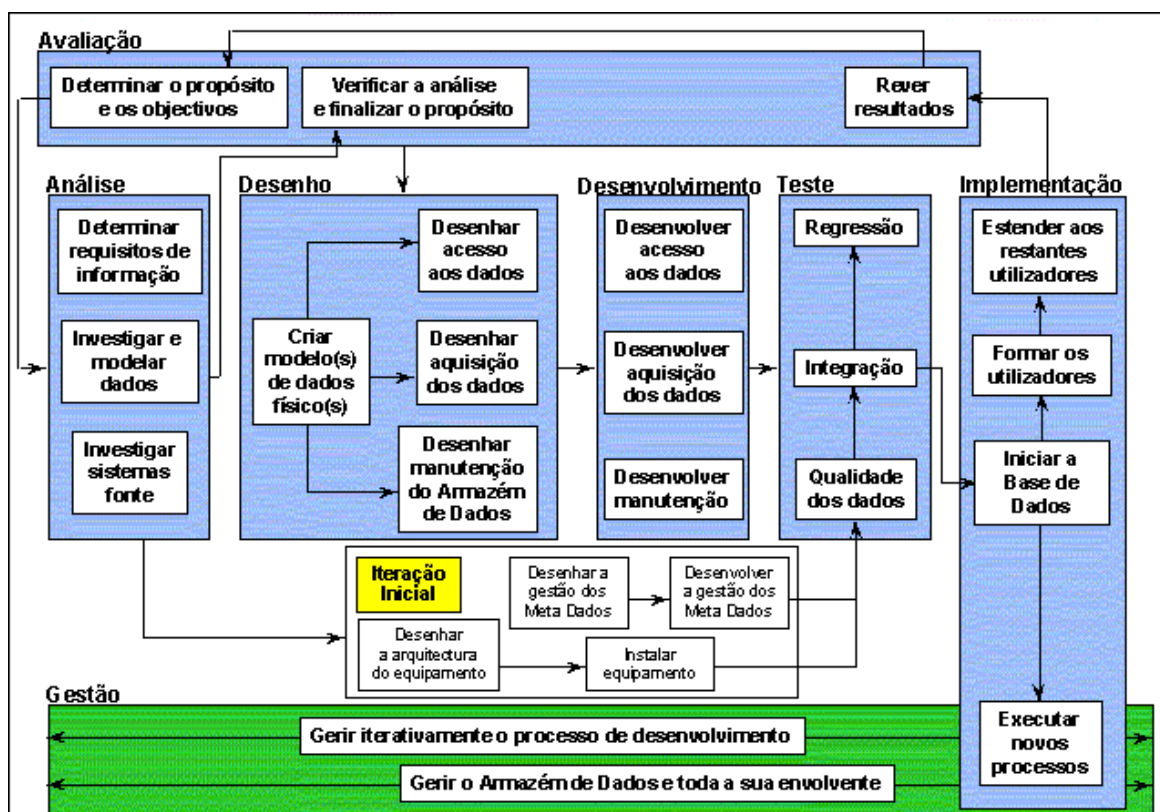


Figura 3.5: Método da DataWing Consulting Services, LLC detalhado (traduzido de [Welch, 1998])

Neste método é dado especial destaque à primeira iteração onde é importante fazer o correcto dimensionamento do equipamento. A definição dos Meta Dados é, também, de particular importância por razões já especificadas (ver subsecção 2.2.2). Apesar de estas tarefas aparecerem como fazendo parte da primeira iteração, a verdade é que estão, de

forma mais ou menos evidente, presentes em todas as iterações; só que na primeira delas, essas tarefas têm uma particular importância.

3.3 Os métodos apresentados face aos factores críticos de sucesso

Dos métodos analisados a nota de maior realce vai para o facto de todos apresentarem uma estrutura cíclica constituída por iterações curtas (máximo de seis meses). Esse facto está directamente relacionado com a grande quantidade de informação envolvida nestes projectos e a necessidade de gerir as expectativas dos utilizadores. É também notório o destaque dado à organização do projecto quer pelas considerações prévias existentes quer pela inclusão de diversos elementos conducentes ao estabelecimento de objectivos claros e mensuráveis, tais como planos de trabalho e especificações de desenvolvimento. Também aqui a gestão das expectativas é acautelada, assim como a boa condução do projecto. A necessidade de compreender os objectivos do negócio está também patente em todos os métodos.

De facto, as diferenças existentes de método para método são mais aparentes do que reais (Figura 3.6). Pode-se, de alguma forma, dizer que as táticas divergem mas que as estratégias são idênticas.

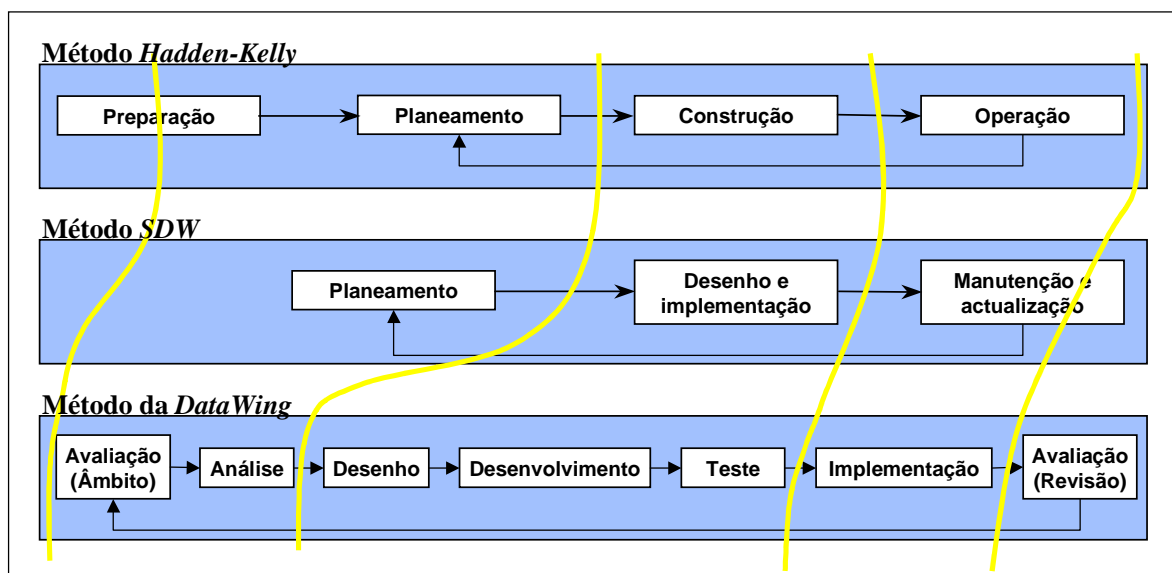


Figura 3.6: Comparação entre os métodos de desenvolvimento

Na Figura 3.7 é apresentado o que se pode designar por método de desenvolvimento genérico tendo como referência a comparação entre os três métodos analisados apresentada na Figura 3.6.

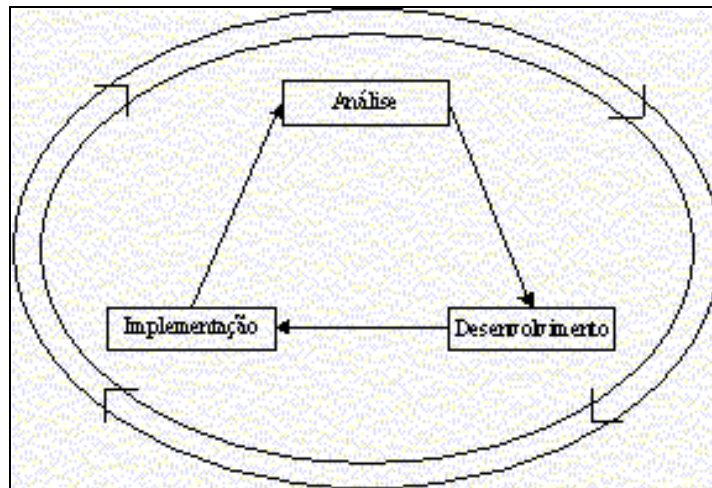


Figura 3.7: Método de desenvolvimento genérico

4 Apresentação do caso de estudo

O caso de estudo considerado tem por objectivo avaliar a aplicabilidade de conceitos de Armazenamento de Dados no desenvolvimento de um módulo de apoio aos decisores no seu processo de tomada de decisões em empresas de transportes colectivos rodoviários de passageiros (Módulo QIO). O desenvolvimento do Módulo QIO (Quadro de Indicadores Operacionais) é efectuado no âmbito do sistema GIST98/EUROBUS, do qual faz parte integrante.

Assim, começa-se por apresentar o consórcio GIST (Gestão Integrada de Sistemas de Transportes). De seguida, é apresentado o sistema GIST98/EUROBUS, sendo dado especial destaque aos dados disponibilizados ou a disponibilizar num futuro próximo pelo referido sistema.

4.1 O consórcio GIST

O consórcio GIST [Sousa et al, 2000] (Figura 4.1) é constituído por cinco empresas de transportes colectivos rodoviários de passageiros e dois institutos de investigação e desenvolvimento.

As cinco empresas de transportes são a CARRIS (Companhia Carris de Ferro de Lisboa, S.A.), a STCP (Sociedade de Transportes Colectivos do Porto, S.A.), os Horários do Funchal (Horários do Funchal Transportes Públicos, Lda.), o Grupo Barraqueiro (Empresa de Camionetes Barraqueiro) e a Vimeca (Vimeca Transportes – Viação Mecânica de Carnaxide, Lda.). Em 1998 estas empresas operavam diariamente com cerca de 6000 viaturas correspondendo a mais de metade das 10151 viaturas do mercado português de transportes colectivos rodoviários de passageiros [INE, 1998].

Os dois institutos são o ICAT (Instituto de Ciências Aplicadas à Tecnologia) e o INEGI (Instituto de Engenharia Mecânica e Gestão Industrial). O primeiro encontra-se associado à Faculdade de Ciências da Universidade de Lisboa enquanto o segundo se encontra associado à Faculdade de Engenharia da Universidade do Porto.

A equipa de desenvolvimento é constituída por elementos dos dois institutos, sendo o trabalho gerido por uma equipa que integra elementos tanto dos institutos como das cinco empresas de transportes que constituem o consórcio GIST.

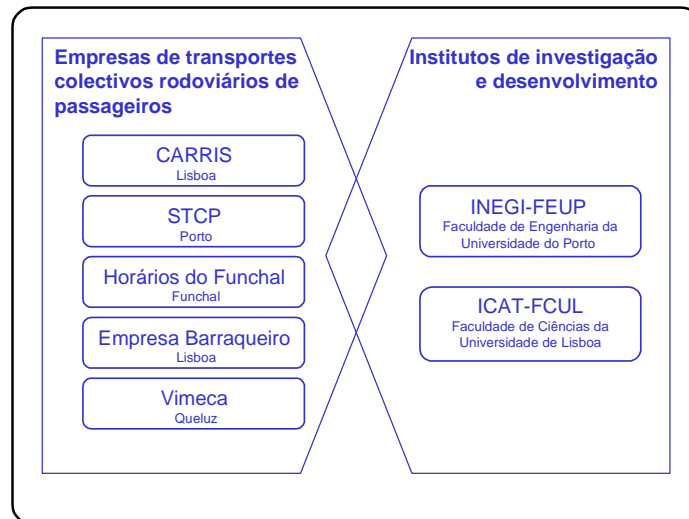


Figura 4.1: O consórcio GIST

4.2 O sistema GIST98/EUROBUS [ICAT & INEGI, 1997]

O sistema GIST98/EUROBUS (Figura 4.2) é um Sistema de Apoio à Decisão para o planeamento de empresas de transportes colectivos rodoviários de passageiros, sendo constituído por onze módulos e três Bases de Dados.

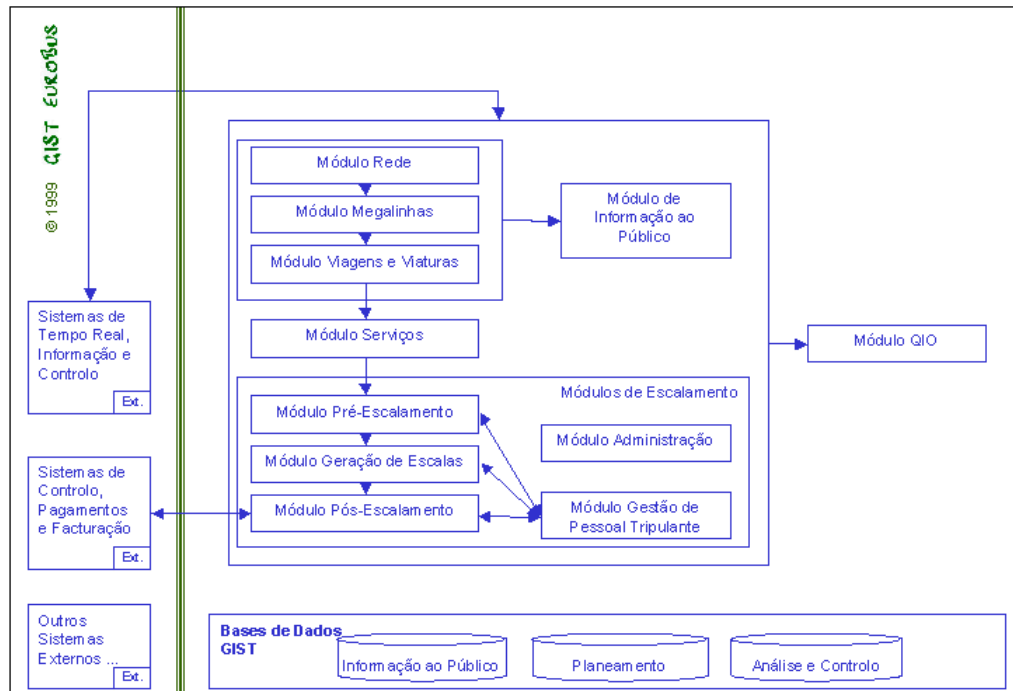


Figura 4.2: Arquitectura de referência do sistema GIST98/EUROBUS

Dada a complexidade do sistema GIST98/EUROBUS, seria fastidioso descrever em detalhe todos os conceitos que lhe são inerentes. No entanto, a descrição deverá dar uma

ideia precisa do potencial de informação dos dados que o compõem, sob risco de não ser possível avaliar eventuais falhas na utilização desses mesmos dados. Na descrição do sistema procura-se um compromisso entre o grau de detalhe e o nível de precisão da informação.

O desenvolvimento do sistema GIST98/EUROBUS é efectuado de acordo com um documento de especificação [ICAT & INEGI, 1997] previamente aprovado pelas empresas.

4.2.1 A definição da rede de transportes: o Módulo Rede

O Módulo Rede (Figura 4.3) tem como função criar e gerir a rede de transportes. É também neste módulo que são definidos alguns dos conceitos necessários para os restantes módulos do sistema.

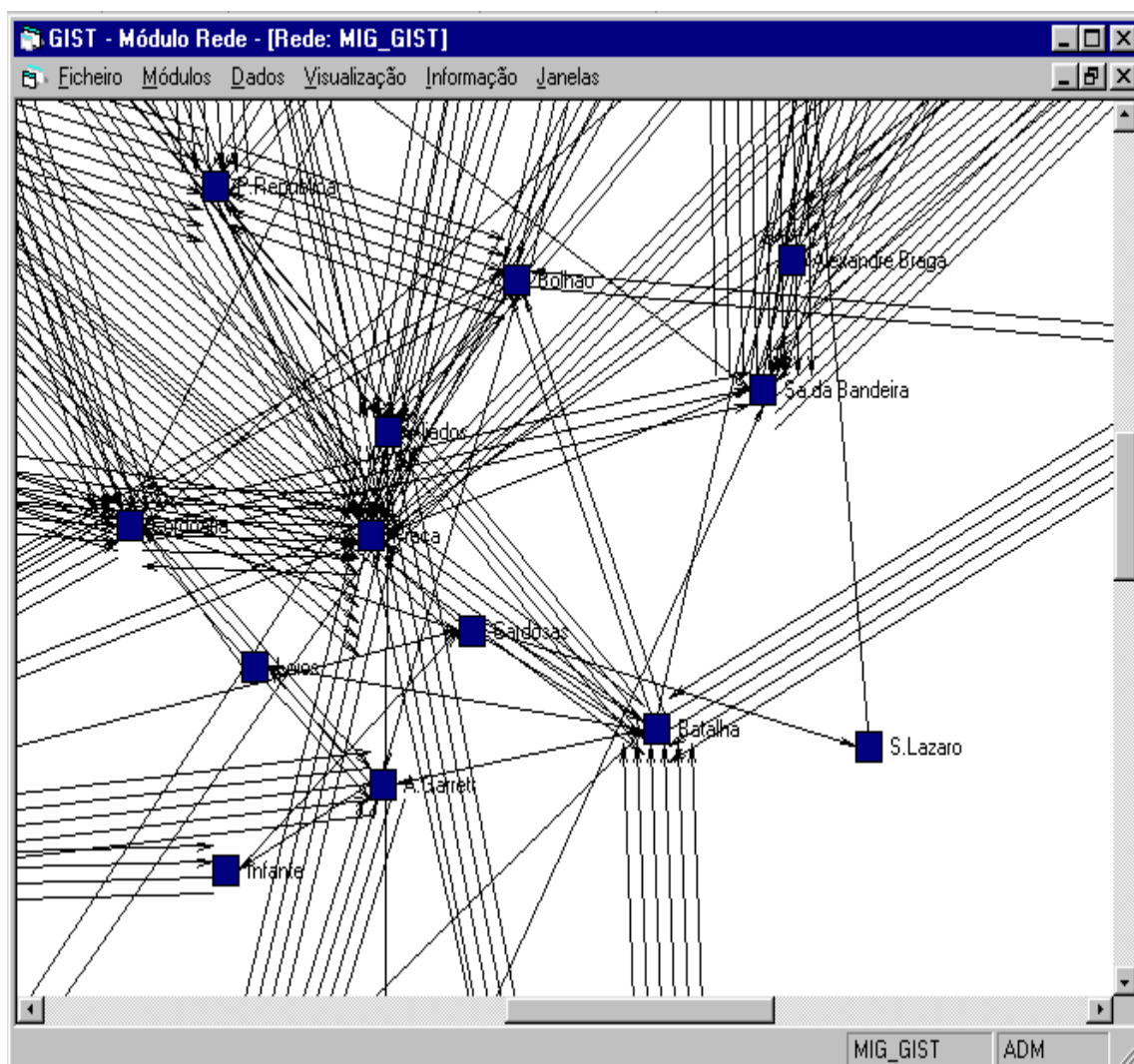


Figura 4.3: Módulo Rede

Os conceitos fundamentais do Módulo Rede são:

- Tipo de viatura, agrupa viaturas com características idênticas para efeitos de planeamento. Exemplos: autocarros articulados, eléctricos.
- Época do ano, identifica um período de tempo mensurável em dias com oferta de viagens diferenciada. Exemplos: período escolar, período de férias, época alta de turismo.
- Tipo de dia, identifica dias com diferentes ofertas de viagens. Exemplos: dias úteis, domingos e feriados.
- Período, é um período de tempo associado a uma época do ano e a um tipo de dia. Identifica períodos de tempo com intensidades de trânsito diferentes. Exemplo: períodos correspondentes às horas de ponta no início da manhã e fim da tarde dos dias úteis da época do ano referente ao período escolar.
- Nó, é um ponto da rede com importância para o planeamento. Exemplos: estações de recolha, paragens de término das viagens. Na Figura 4.3 ‘S. Lazaro’ e ‘Batalha’ são exemplos da representação gráfica de nós.
- Segmento, é um trajecto directo entre dois nós sem passar por nenhum nó intermédio. Um segmento tem a orientação definida, ou seja, caso se pretenda definir um trajecto entre dois nós com circulação em ambos os sentidos, é necessário criar dois segmentos com orientações contrárias. O tempo de duração associado a um segmento pode ser diferente de período para período. Uma vez que um período se refere a uma época do ano e a um tipo de dia, implicitamente a duração associada ao segmento pode variar em função do período do dia, do tipo de dia e da época do ano. O segmento de recta orientado entre os nós ‘Cardosas’ e ‘S. Lazaro’ na Figura 4.3 é um exemplo da representação gráfica de um segmento.

4.2.2 A definição das linhas: o Módulo Megalinhas

O Módulo Megalinhas permite definir o conceito de linha (ou carreira) que constitui o elemento básico de identificação, por parte dos utentes, da oferta de viagens das empresas⁴.

Os conceitos fundamentais do Módulo Megalinhas são:

⁴ As linhas (ou carreiras) são habitualmente identificadas por números nos meios urbanos, enquanto nos meios suburbanos são identificadas pelos locais de término (tipicamente o nome de duas localidades).

- Percurso, é definido por uma sequência ordenada de segmentos conexos⁵, daí resultando o facto de os percursos terem uma orientação implícita. É sobre os percursos que são definidas as viagens.
- Linha, é um conjunto de percursos. Tipicamente, uma linha é constituída por dois percursos de sentidos contrários sendo, no entanto, comum existirem outros percursos alternativos que são utilizados apenas em determinados tipos de dia ou épocas do ano.
- Percurso exterior, é um percurso que não pertence a nenhuma linha. Tipicamente são percursos sobre os quais são definidas viagens que não levam passageiros. Exemplo: percursos de ligação entre as estações de recolha e as linhas.

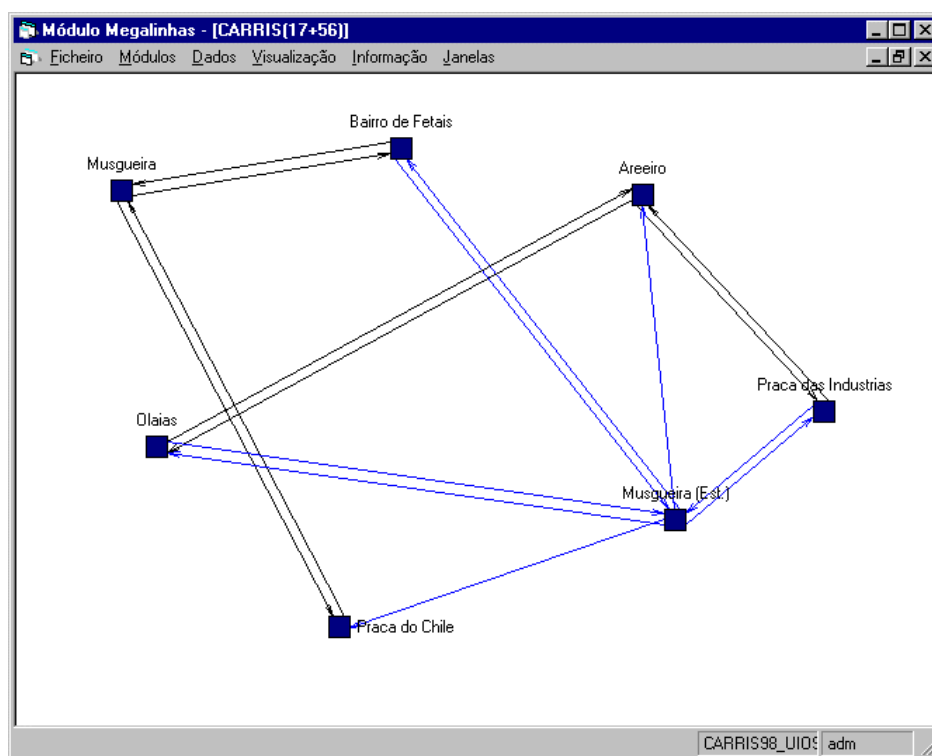


Figura 4.4: Módulo Megalinhhas

- Megalinha, é um conceito exclusivo do sistema GIST98/EUROBUS sendo a sua unidade básica de planeamento. Uma megalinha (Figura 4.4) é definida como sendo um conjunto de linhas e de percursos não pertencentes a essas linhas. A utilização deste conceito surgiu da necessidade de efectuar o planeamento quer das viaturas quer dos tripulantes para mais do que uma linha em simultâneo, dado que o planeamento linha a linha acarreta para a maioria das empresas custos mais elevados. A megalinha surge

⁵ O nó de fim de um segmento deve ser o nó de início do segmento seguinte.

como um compromisso entre o custo das soluções obtidas e o grau de complexidade dos problemas.

- Estatuto de um nó, é uma identificação dada a nós especiais. Existem por defeito dois estatutos de nó: estação de recolha, que identifica um parque de viaturas; e ponto de rendição, que identifica um local onde um tripulante pode entrar ou sair de serviço.

4.2.3 A definição das viagens e do serviço das viaturas: o Módulo Viagens e Viaturas

O Módulo Viagens e Viaturas (Figura 4.5) tem duas funcionalidades essenciais: a definição da oferta de viagens ao público e a definição do serviço das viaturas, i.e., para uma determinada megalinha são definidas as viagens a realizar para um certo tipo de dia e época do ano e, posteriormente, especificadas as viagens que cada viatura deverá realizar. É possível recorrer a diversas heurísticas [Paixão, 1987], [Branco, 1989], [Mesquita & Paixão, 1990] para a definição dos serviços das viaturas.

Os conceitos fundamentais do Módulo Viagens e Viaturas são:

- Viagem, é definida sobre um percurso a uma determinada hora do dia. Podem ser de dois tipos: ao público ou em vazio. As viagens ao público pressupõem o transporte de passageiros enquanto as viagens em vazio podem, eventualmente, levar passageiros não fazendo, no entanto, parte dos horários disponibilizados aos utentes potenciais. Na Figura 4.5 cada segmento de recta representa uma viagem sendo o tempo apresentado no eixo horizontal e os nós de passagem das viagens no eixo vertical.
- Especificação de viagem, é um conceito que facilita a criação sistemática de viagens, nomeadamente através da definição de atributos que serão comuns a um conjunto de viagens a criar. As viagens podem ser singulares ou periódicas. No primeiro caso, a especificação serve de base à criação de uma só viagem enquanto que, no segundo, permite criar um conjunto de viagens.
- Turno, é definido como sendo o conjunto de viagens a realizar por uma viatura num dia de trabalho. O sistema GIST98/EUROBUS não tem informação sobre as viaturas físicas, tem apenas a informação do serviço a realizar por cada viatura. Exemplo: uma viatura sai às 6:30 da estação de recolha ficando de serviço na linha 20 até às 22:00. Esta informação é dada pelo sistema, nomeadamente todas as viagens que constituem o turno. Qual a viatura física que realiza o turno não é possível saber através deste sistema. Os turnos podem ser definidos manualmente ou recorrendo a heurísticas.

- Horário de viaturas, é o conjunto de turnos criados com base em viagens previamente definidas para uma megalinha, um tipo de dia e uma época do ano.

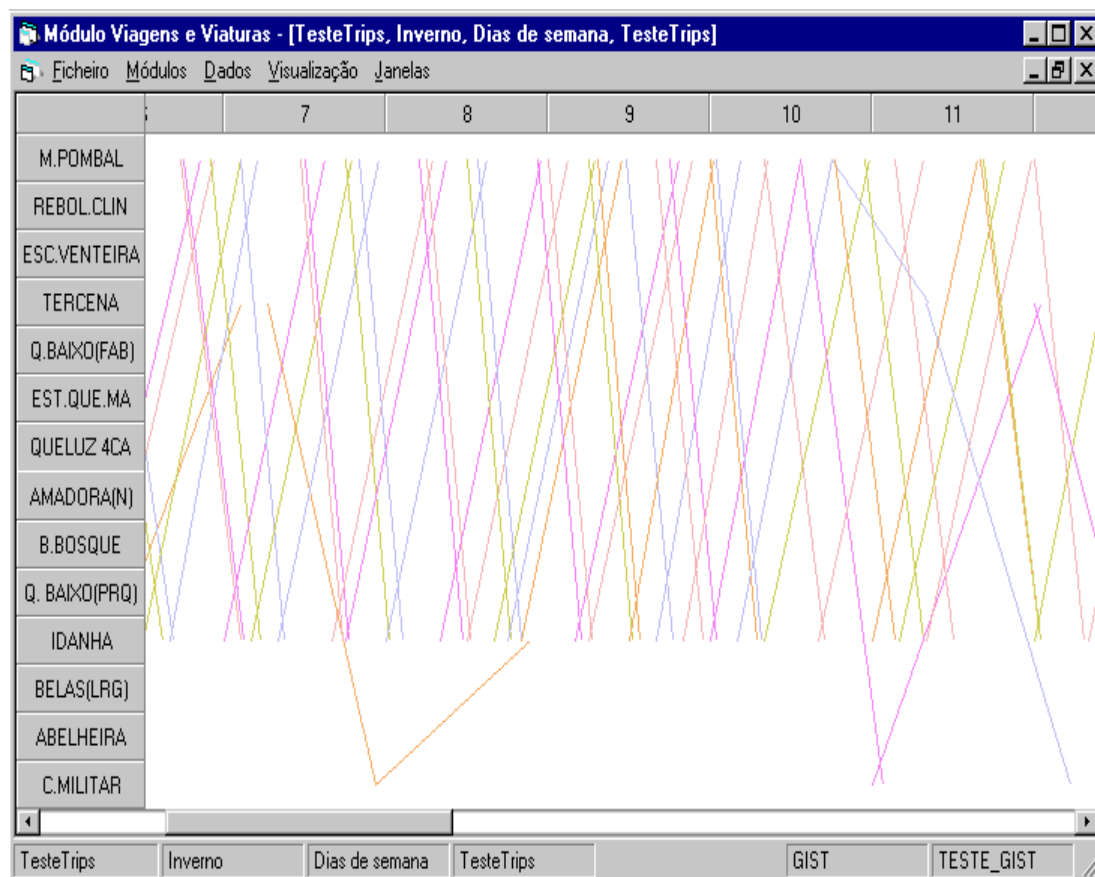


Figura 4.5: Módulo Viagens e Viaturas

4.2.4 A definição dos serviços a realizar e a sua afectação aos tripulantes: os módulos de Serviços e de Escalamento⁶

A finalidade do conjunto dos módulos que aqui se descrevem é o de definir os serviços a realizar, afectá-los aos tripulantes para cada dia e registar o serviço efectivamente realizado permitindo, assim, efectuar o processamento dos salários. Todo este processo é realizado recorrendo a seis módulos que se passam a descrever.

O Módulo Serviços

O Módulo Serviços permite a definição dos serviços a realizar pelos tripulantes tendo como base os turnos criados no Módulo Viagens e Viaturas. Deve contribuir para uma gestão optimizada desses serviços garantindo as regras laborais em vigor na empresa. Para

⁶ O termo correcto é escalonamento, no entanto, as empresas utilizam o termo escalamento (de escala de trabalho) sendo também esse o nome dado ao conjunto dos módulos do sistema GIST98/EUROBUS que realizam essa tarefa [Paes, 1999].

tal, é possível recorrer a diversas heurísticas de optimização [Paixão & Pato, 1989], [Dias, 1995], [Portugal, 1998], [Couto, 1998].

Os conceitos fundamentais do Módulo Serviços são:

- Tramo, é um intervalo de tempo entre duas passagens consecutivas de uma viatura por pontos de rendição. A cada tramo está associado um conjunto de viagens. Na Figura 4.6 cada rectângulo representa um tramo.
- Serviço, é o trabalho a realizar por um tripulante definindo-se como uma sequência de tramos sujeita a um conjunto de regras. Um serviço pode ser definido manualmente e/ou recorrendo a heurísticas de optimização. Na Figura 4.6 o conjunto de tramos com a mesma identificação (s1, s2, etc.) representa um serviço.
- Etapa, é um período de tempo contínuo durante o qual os tripulantes estão ao serviço da empresa. Um serviço pode ter uma ou mais etapas.
- Folga, é o intervalo de tempo entre duas etapas durante o qual os tripulantes não se encontram ao serviço da empresa.
- Pequena interrupção de serviço, é um intervalo de tempo entre dois períodos de condução (ou tramos), em que os tripulantes se mantêm ao serviço da empresa.
- Tipo de serviço, é um padrão de serviço, ou seja, um conjunto de regras que todos os serviços desse tipo devem respeitar. Cada serviço só pode estar associado a um tipo de serviço. Os tipos de serviço devem garantir o cumprimento das regras laborais em vigor na empresa.
- Horário de serviços (Figura 4.6), é o conjunto de serviços criados com base em um ou mais horários de viaturas relativos aos mesmos tipo de dia e época do ano.

O Módulo Administração

Permite a definição de conceitos gerais aos Módulos de Escalamento [Ferreira, 1988], [Borges, 1994], [Ferreira et al, 1998]. É um módulo de configuração onde, entre outras funcionalidades, são especificados os parâmetros que permitem adaptar o processo de escalamento à realidade de cada empresa.

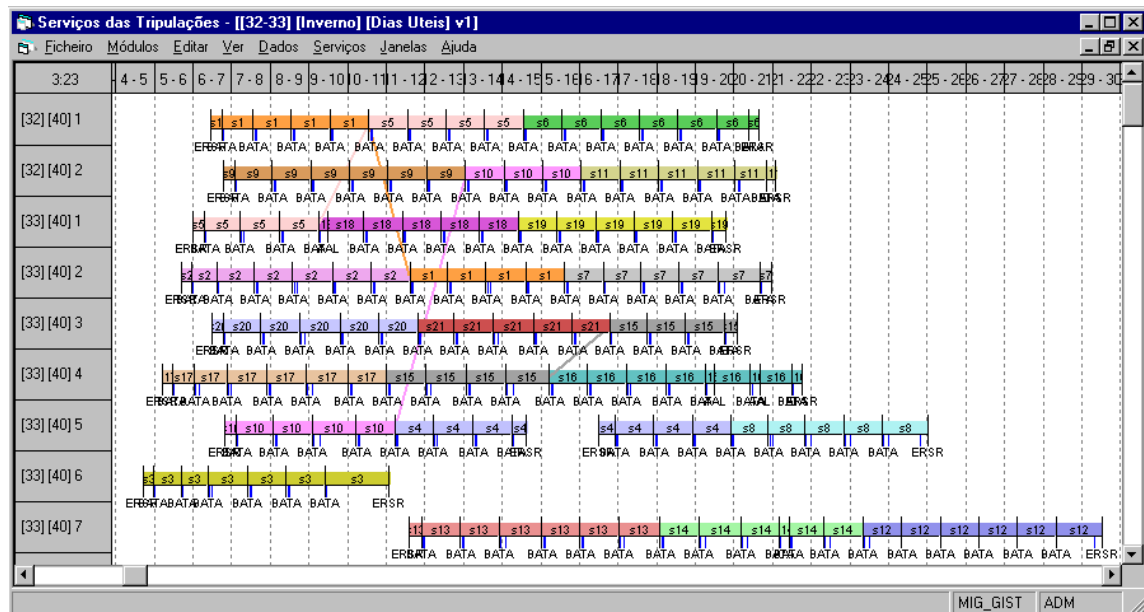


Figura 4.6: Módulo Serviços

O Módulo Gestão de Pessoal Tripulante

O Módulo Gestão de Pessoal Tripulante [Pedrosa et al, 2000] permite gerir toda a informação relativa aos tripulantes (Figura 4.7), nomeadamente características pessoais, histórico, indisponibilidades (exemplos: férias, baixa) e pedidos dos tripulantes tais como mudança de estação de recolha ou outros.

Nº de Ordem	Nº de Chapa	Nome	Apelido	Data Situação...	Acti...	Dat
041670	0006060	AMERICO LOUREIRO DOS	SANTOS	01-10-1997	X	
046981	0006105	FELICIANO DA SILVA	QUARESMA	01-07-1998	X	
048836	0004350	JOAO TAVARES	LUCAS	01-10-1997	X	
048941	0004018	LUIS DA COSTA	GERMANO	01-10-1997	X	
049450	0004059	DANIEL LOBATO	MARQUES	01-10-1997	X	
051900	0006357	ANTONIO SALGUEIRO	CUNHA	01-10-1997	X	
052540	0004381	JOAO DIAS	NUNES	01-10-1997	X	
054291	0004050	FRANCISCO MARTINS DAS	NEVES	01-10-1997	X	
055611	0004365	JOSE MIRANDA	MEIO-TOSTAO	01-10-1997	X	
056871	0006449	JOSE VITERBO PRATA	GARRIDO	01-10-1997	X	
057126	0006062	JOSE DE OLIVEIRA	CORREIA	01-07-1998	X	
057291	0006016	JOAO DIOGO	PIRES	01-10-1997	X	
057436	0004290	MANUEL CARDOSO RESENDE	JORGE	01-10-1997	X	
057541	0004385	MANUEL DA SILVA	COVELINHAS	01-10-1997	X	
057568	0004043	JOSE DE JESUS	JOAO	01-10-1997	X	
057827	0004425	CONSTANTINO DA SILVA	DIOGO	01-10-1997	X	
058416	0004062	JOSE CRISANTE	LOURO	01-10-1997	X	
058912	0004380	JOAQUIM GAMA	LOBATO	01-10-1997	X	
059889	0006108	SEBASTIAO LARANJEIRA DA ...	LOURENCO	01-10-1997	X	

Figura 4.7: Módulo Gestão de Pessoal Tripulante

Os conceitos fundamentais do Módulo Gestão de Pessoal Tripulante são:

- Tipo de situação, é a designação dada a um estado de um tripulante. Exemplos: baixa, férias, falta injustificada.
- Situação, é a afectação de um tipo de situação a um período de tempo (em dias).
- Categoria, é a designação da posição ocupada pelo tripulante dentro da empresa. Exemplos: motorista, guarda freio, técnico auxiliar.
- Grupo de folga, determina a regra de atribuição da folga semanal ao tripulante. Exemplos: Sábado, Domingo, rotativa.
- Grupo de férias, permite atribuir períodos de férias aos tripulantes de forma a garantir o serviço mínimo por parte da empresa. Por exemplo, permite assegurar que não vão de férias em Agosto todos os tripulantes.

O Módulo Pré-Escalamento

O Módulo Pré-Escalamento [Pedrosa et al, 2000] permite definir as escalas modelo que servirão de base à definição do trabalho a efectuar diariamente por cada tripulante. Usa como base de trabalho os serviços criados no Módulo Serviços.

Os conceitos fundamentais do Módulo Pré-Escalamento são:

- Serviço de escala, pode ter origem nos serviços criados no Módulo Serviços ou ser criado no próprio Módulo Pré-Escalamento. Neste último caso, corresponde normalmente a outras tarefas que não o da realização de viagens ao público, tais como manobras, reservas, etc. Pode acontecer que alguns serviços de escala não tenham viagens associadas apesar de, na prática, o serviço incluir a realização de viagens. Isto deve-se à possibilidade de criar serviços de escala novos ou alterar outros já existentes sem ser necessário especificar as viagens associadas para, assim, simplificar essa tarefa.
- Situação na escala, estabelece diferentes formas de atribuição dos tripulantes aos serviços. Exemplos: efectivo, substituto, supra.
- Grupo de serviços, é a unidade básica de planeamento dos Módulos de Escalamento sendo constituído por um conjunto de serviços de escala e um conjunto de tripulantes. A um grupo de serviços está associado um esquema de rotação, ou seja, as regras de atribuição dos serviços de escala aos tripulantes. Na Figura 4.8 a cada linha corresponderá um tripulante. Assim, na primeira semana o tripulante da posição 1 está

de folga no Domingo faz o serviço de escala nº 2 de Segunda a Sexta Feira e está de novo de folga no Sábado.

Afectação de tripulantes ao grupo Grupo 1/83

Esquema de Rotação | Serviços | Tripulantes

Semana 1

	S	M	T	W	T	F	S
1	F	2	2	2	2	2	F
2	F	F	3	3	3	3	3
3	2	F	F	4	4	4	4
4	3	3	F	F	5	5	5
5	4	4	4	F	F	1	1
6	5	5	5	5	F	F	2
7	1	1	1	1	1	F	F

Semana 2

	S	M	T	W	T	F	S
F	F	4	4	4	4	4	
3	F	F	5	5	5	5	
4	4	F	F	1	1	1	
5	5	5	F	F	2	2	
1	1	1	1	F	F	3	
2	2	2	2	2	F	F	
F	3	3	3	3	3	F	

Semana 3

	S	M	T	W	T	F	S
4	F	F	1	1	1	1	
5	5	F	F	2	2	2	
1	1	1	F	F	3	3	
2	2	2	2	F	F	4	
3	3	3	3	3	F	F	
F	4	4	4	4	4	F	
F	F	5	5	5	5	5	

Semana 4

	S	M	T	W	T	F	S
1	1	1	F	F	3	3	3
2	2	2	2	F	F	4	4
3	3	3	3	3	F	F	5
4	4	4	4	4	4	F	F
5	F	5	5	5	5	5	F
6	F	F	1	1	1	1	1
7	5	F	F	2	2	2	2

Semana 5

	S	M	T	W	T	F	S
3	3	3	F	F	5	5	
4	4	4	4	F	F	1	
5	5	5	5	5	5	F	
F	1	1	1	1	1	F	
F	F	2	2	2	2	2	
1	F	F	3	3	3	3	
2	2	F	F	4	4	4	

Semana 6

	S	M	T	W	T	F	S
5	5	5	5	F	F	2	
1	1	1	1	1	F	F	
F	2	2	2	2	2	F	
F	F	3	3	3	3	3	
2	F	F	4	4	4	4	
3	3	F	F	5	5	5	
4	4	4	F	F	1	1	

Afectação:

	Tripulante Efectivo	Tripulante Substituto
1		
2		
3		
4		
5		

OK Cancelar Limpar Imprimir Ajuda

Figura 4.8: Módulo Pré-Escalamento

- Escala modelo, é o conjunto dos grupos de serviço de uma estação de recolha que servirá de base à geração das escalas diárias durante um período de tempo especificado.
- Equivalência, é um conceito determinante para a posterior atribuição dos serviços diários aos tripulantes. De facto, a atribuição do serviço ao tripulante está relacionado com a sua posição relativa no grupo a que pertence. Se o dia em que o tripulante ocupa essa posição for um Domingo, por exemplo, o serviço atribuído será diferente daquele que seria atribuído se fosse um dia útil, ou seja, há serviços com características

idênticas correspondentes a diferentes ofertas de viagens ao público. A equivalência estabelece quais os serviços equivalentes para diferentes ofertas de viagens ao público.

O Módulo Geração de Escalas

O Módulo Geração de Escalas [Pedrosa et al, 2000] permite definir o trabalho a realizar por cada tripulante para um dia específico usando as escalas modelo como base de trabalho.

Os conceitos fundamentais do Módulo Geração de Escalas são:

- Serviço diário, é o serviço a realizar por um tripulante num determinado dia. Sendo criado a partir de um serviço de escala pode, no entanto, ser alterado permitindo assim a flexibilização do processo de geração de escalas diárias.
- Linha de escala, tem a afectação dos tripulantes aos serviços diários, ou seja, é na linha de escala que é definido quem vai fazer que serviço para um determinado dia. A cada serviço é atribuído um motorista efectivo ou um suplente e, eventualmente, um supra. Os motoristas suplentes são os substitutos dos motoristas efectivos quando estes se ausentam por um período pré-anunciado. Exemplo: baixa por doença. Os motoristas supras são os que substituem os motoristas efectivos ou suplentes em situações pontuais.
- Escala diária (Figura 4.9), é o conjunto de serviços diários para um determinado dia a realizar pelos tripulantes de uma estação de recolha.
- Escala afixada, é a versão definitiva da escala diária, ou seja, aquela que é tornada pública.

O Módulo Pós-Escalamento

O Módulo Pós-Escalamento permite registar as diferenças entre a escala afixada e a escala efectivamente realizada. Permite fornecer os dados necessários ao processamento de salários do pessoal tripulante. Usa como base de trabalho as escalas afixadas.

Os conceitos fundamentais do Módulo Pós-Escalamento são:

- Escala real, é o conjunto de serviços realizados num determinado dia pelos tripulantes de uma estação de recolha. A escala real é obtida pelo registo das alterações dos serviços realizados face aos afixados.

- Situação real, estabelece o tipo de situação efectivamente verificado num dia por um tripulante.

Grupo	Posição	Serviço	Duração	Efectivo	Situação	Substituto	Situação	S
Grupo 1/83	1	5	08:06	0003001	SSS			
	2			0003002	FFF			
	3			0003003	FFF			
	4	1	06:57	0003004	SSS			
	5	2	07:37	0003005	SSS			
	6	3	07:45	0003006	SSS			
	7	4	07:37	0003007	SSS			
Grupo 2/83	1	10	07:21	0003008	SSS			
	2			0003009	FFF			
	3			0003010	FFF			
	4	6	07:43	0003011	SSS			
Grupo 1/1	1	11	06:08	0003015	SSS			
	2	12	07:56	0003016	SSS			

Nº de Cha...	Nome	Apelido	H. F. Dia Anterior	Horas por Asseg...
0003001	PAULO MANUEL SOUSA	HEITOR		00:00
0003002	ABEL DO NASCIMENTO	ESCOBAR		00:00
0003003	JOAQUIM FARINHA DA	SILVA		00:00
0003004	PEDRO PAULO MORGADINHO	EMPADINHAS		00:00
0003005	DOMINGOS CORREIA	BORREGO		00:00
0003006	FERNANDO DOS SANTOS	NUNES		00:00
0003007	DINIS MARIA	MARTINS		00:00
0003008	ARNALDO VAZ DA	CRUZ		00:00
0003009	ANTONIO JACINTO FRANCIS...	PARDAL		00:00
0003010	CARLOS ALBERTO GUEDES	MORAIS		00:00

Figura 4.9: Módulo Geração de Escalas

4.2.5 A disponibilização de informação ao público: o Módulo de Informação ao Público

O Módulo de Informação ao Público tem três componentes: a primeira tem como utilizadores elementos da empresa de transportes e permite a definição das paragens; a segunda tem como utilizadores elementos da empresa de transportes (não necessariamente os mesmos da componente anterior) e permite a extracção de informação sobre horários de paragens para a criação de documentos; a terceira componente tem como utilizadores alvo os potenciais utentes dos serviços da empresa e permite a obtenção pela *internet* de informação sobre as linhas, paragens e respectivos horários, assim como a identificação do melhor percurso entre dois pontos segundo um critério à escolha. Esta componente, designada por INFOBUS [Barros et al, 1999] (Figura 4.10), utiliza um sistema de informação geográfica para a representação das linhas sobre um mapa da área onde a empresa opera.

Os conceitos fundamentais do Módulo de Informação ao Público são:

- Local de referência, é um local facilmente identificável pelo público, servindo como ponto de referência de uma determinada área.
- Local de interesse, é um local com interesse turístico. Os locais de interesse são também locais de referência.
- Paragem, é um local de entrada e/ou saída de passageiros.
- Linha, é, tal como o conceito linha definido no Módulo Megalinhas, o elemento básico de identificação, por parte dos utentes, da oferta de viagens das empresas. No entanto, no Módulo de Informação ao Público, o conceito linha compreende a sua representação gráfica sobre um mapa.

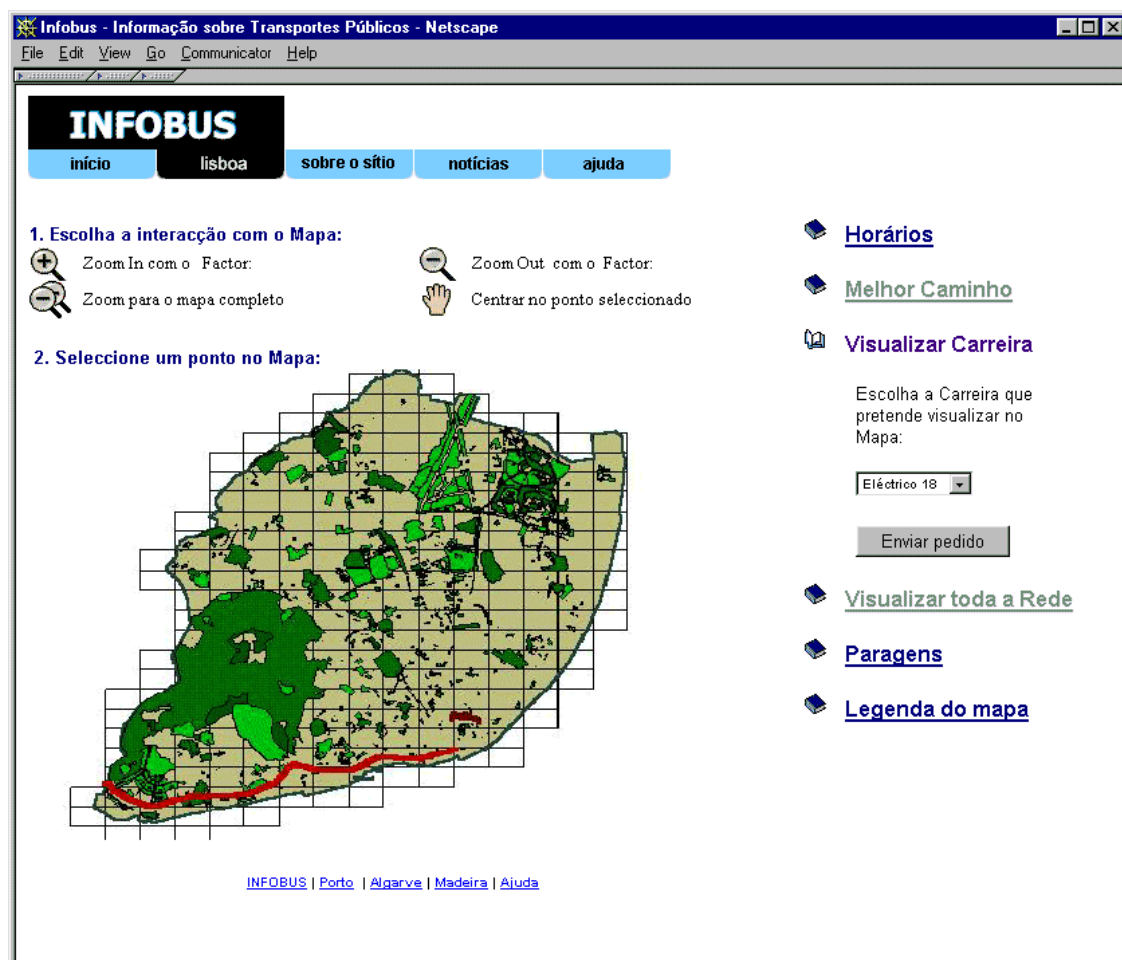


Figura 4.10: Componente INFOBUS do Módulo de Informação ao Público

4.2.6 A disponibilização de informação aos agentes de decisão: o Módulo QIO

O Módulo QIO (Quadro de Indicadores Operacionais) tem como objectivo disponibilizar informação de planeamento e controlo aos agentes de decisão das empresas de transportes,

para os apoiar no processo de tomada de decisões. Essa informação deve ser obtida exclusivamente a partir do sistema GIST98/EUROBUS, uma vez que é a única fonte de dados comum a todas as empresas do consórcio GIST.

A concepção, desenvolvimento e implementação do Módulo QIO como um sistema de Armazenamento de Dados é o caso de estudo desta dissertação.

4.2.7 O conceito de versão activa

O conceito de versão activa é transversal a vários módulos do sistema GIST98/EUROBUS. Designa-se por versão activa uma versão que entra em vigor por um determinado período de tempo. Este conceito aplica-se a horários de viaturas (subsecção 4.2.3), horários de serviços, escalas modelo e escalas diárias (subsecção 4.2.4). O conceito de escala afixada (subsecção 4.2.4) sobrepõe-se ao conceito de versão activa sendo utilizado em vez deste último, por forma a respeitar a cultura própria das empresas do consórcio. O conceito de versão activa é fundamental para o Módulo QIO pois só essas versões é que são relevantes para a obtenção de indicadores. As versões que não estão activas são, tipicamente, versões de trabalho usadas para o estudo de diferentes cenários.

4.2.8 As Bases de Dados do sistema

O sistema GIST98/EUROBUS tem três Bases de Dados:

- Base de Dados de planeamento, é a Base de Dados operacional do sistema. Contém todos os dados do sistema com excepção dos existentes na Base de Dados de informação ao público e dos dados necessários ao Módulo QIO. Uma mesma empresa pode ter mais do que uma Base de Dados de planeamento, dado que só é possível definir uma rede de transportes para cada Base de Dados de planeamento. Assim, se uma mesma empresa tiver mais do que um tipo de rede, então deverá ter uma Base de Dados de planeamento por cada um deles, por exemplo, rede de eléctricos ou rede de autocarros. Note-se que o conceito de tipo de rede pressupõe uma rede física diferente. A rede de eléctricos e de autocarros são fisicamente distintas.
- Base de Dados de informação ao público, tem a informação geográfica necessária ao Módulo de Informação ao Público, em particular à sua componente INFOBUS.
- Base de Dados de análise e controlo, contém os dados necessários ao Módulo QIO (ver secção 5.1).

5 O Módulo QIO face aos sistemas de Armazenamento de Dados

Neste capítulo é feito o enquadramento do Módulo QIO face aos sistemas de Armazenamento de Dados, nomeadamente quanto à pertinência da utilização desta tecnologia na concepção do módulo. De seguida, é analisada a aplicabilidade dos métodos de desenvolvimento previamente analisados (capítulo 3) para o caso em estudo.

5.1 A construção de um Armazém de Dados no âmbito do sistema GIST98/EUROBUS

Os dados operacionais disponíveis, em particular os da Base de Dados de planeamento, são:

- Organizados em função dos dados, ou seja, a organização dos dados é feita de acordo com os próprios dados (Bases de Dados relacionais na 3ª forma normal) e não com o fim a que se destinam, resultando em pouca flexibilidade na leitura dos dados por parte do utilizador final e em tempos de resposta às questões por ele colocadas inadequados para o fim a que se destinam;
- Referentes ao momento actual, não reflectindo a evolução ao longo do tempo, i.e., não se pode fazer uma análise temporal de grande parte da informação, dado que só é mantida a versão em uso. Constitui excepção toda a informação sobre o pessoal tripulante, já que é mantido o histórico tanto do trabalho planeado como do realizado.

Assim, surge a necessidade de criar um repositório de dados próprio para o Módulo QIO que garanta que os dados são temáticos, não voláteis e com o histórico da evolução ao longo do tempo. Só assim será possível efectuar, por exemplo, análises temporais adequadas para o apoio à decisão.

Uma vez que a informação deve ser obtida exclusivamente a partir do sistema GIST98/EUROBUS (subsecção 4.2.6), pode-se concluir que o repositório de dados a construir terá as características de um Armazém de Dados Departamental independente (subsecção 2.1.4). De facto, a informação constante no sistema diz respeito ao planeamento e controlo dos serviços das viaturas e dos tripulantes, informação essa que faz parte de uma área funcional específica nas empresas de transportes colectivos de passageiros. No entanto, uma vez que o sistema GIST98/EUROBUS não cobre toda a área funcional

referida, nomeadamente o controlo do serviço das viaturas, o Armazém de Dados Departamental a desenvolver será forçosamente incompleto.

5.2 A especificidade do método de desenvolvimento para o caso em estudo

O método de desenvolvimento adoptado (Figura 5.1) na construção do Módulo QIO surge na sequência do estudo de métodos de desenvolvimento de projectos de Armazenamento de Dados previamente apresentado (capítulo 3). As etapas propostas resultam do método de desenvolvimento genérico apresentado na Figura 3.7. As fases que constituem cada uma destas etapas estão presentes em cada um dos três métodos estudados de acordo com a correspondência existente entre esses métodos e o método de desenvolvimento genérico apresentado na Figura 3.6.

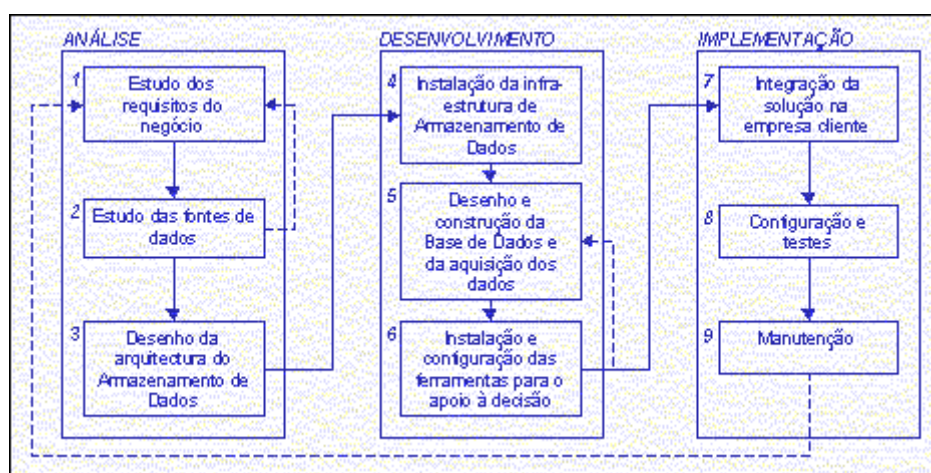


Figura 5.1: Método de desenvolvimento do Módulo QIO

O método de desenvolvimento do Módulo QIO é iterativo sendo composto por três etapas:

- **Análise:** compreende todo o trabalho prévio ao desenvolvimento. É constituída pela fases seguintes:
 - ❑ **Estudo dos requisitos do negócio:** levantamento da informação necessária ao processo de tomada de decisões. Esta fase requer uma grande disponibilidade da parte de diversos elementos da empresa cliente, em particular os agentes de decisão.
 - ❑ **Estudo das fontes de dados:** análise das fontes de dados existentes na empresa cliente e estudo da forma de obtenção da informação pretendida com os dados existentes.

- ❑ Desenho da arquitectura do Armazenamento de Dados: selecção do *software* e do *hardware* e definição da arquitectura do Armazenamento de Dados.
- Desenvolvimento: é a etapa menos visível para o cliente sendo, no entanto, aquela que mais tempo demora. É constituída pelas seguintes fases:
 - ❑ Instalação da infra-estrutura de Armazenamento de Dados: instalação e configuração do *hardware* e *software* seleccionados na fase anterior.
 - ❑ Desenho e construção da Base de Dados e da aquisição de dados: é também nesta fase que é definido o repositório de Meta Dados..
 - ❑ Instalação e configuração das ferramentas para o apoio à decisão: instalação e configuração da ferramenta OLAP e das ferramentas a disponibilizar aos clientes finais.
- Implementação: corresponde à colocação do trabalho desenvolvido na empresa cliente. Compreende as seguintes fases:
 - ❑ Integração da solução na empresa cliente: instalação do *hardware* e *software* necessários.
 - ❑ Configuração e testes: o sistema está em condições de ser utilizado pelos clientes finais quando esta fase estiver concluída desde que, naturalmente, existam dados disponíveis nas Bases de Dados fonte.
 - ❑ Manutenção: vistoria periódica ao Armazém de Dados para correcção de eventuais erros e reajuste na configuração das ferramentas.

Uma vez que o trabalho apresentado é académico alguns dos factores críticos discutidos (secção 3.1) não são aplicáveis ao caso presente, nomeadamente quanto às expectativas em termos de prazos de execução criadas por parte do cliente (as empresas de transportes do consórcio GIST). Assim, as iterações com duração de três a seis meses deixam de ser uma prioridade, ficando desde logo eliminada grande parte da dificuldade processual, sendo o problema centrado na sua componente técnica.

Dado que os conhecimentos prévios sobre a tecnologia e o processo de Armazenamento de Dados eram nulos, o método de desenvolvimento adoptado para o caso em estudo reflecte, necessariamente, esse facto. Assim, a organização sequencial patente na Figura 5.1 foi por diversas vezes interrompida com saltos a fases anteriores e o consequente retomar da

ordem estabelecida no método. Em particular, no desenho e construção da Base de Dados e na aquisição dos dados a experiência é determinante, pois o desenho da Base de Dados requer um bom conhecimento das ferramentas para o apoio à decisão, e a aquisição dos dados é fortemente dependente desse desenho. Estas duas fases foram realizadas quase em simultâneo com base no método de prototipificação evolutiva [Nunes & Cunha, 1998].

Ainda uma nota para o impacto da integração do Armazém de Dados no sistema GIST98/EUROBUS no método de desenvolvimento, e a consequente limitação das fontes de dados. O estudo dos requisitos do negócio é fortemente condicionado por essa limitação sendo as duas primeiras fases (estudo dos requisitos do negócio e estudo das fontes de dados) realizadas em simultâneo. Em vez de se estudar como obter os dados sabendo à partida qual a informação pretendida, a abordagem adoptada aproxima-se mais de responder à pergunta: “qual a informação que se pode obter com os dados disponíveis?”.

6 Aplicação do método de desenvolvimento

O trabalho apresentado foi realizado maioritariamente na empresa STCP, sem que se tivessem descurado aspectos particulares a cada uma das restantes empresas do consórcio GIST.

São descritas apenas as duas primeiras etapas (análise e desenvolvimento) da primeira iteração do método de desenvolvimento utilizado (secção 5.2), sendo a sua apresentação realizada segundo a ordem estabelecida no método. A etapa de implementação não foi realizada devido a atrasos na implementação do Sistema GIST98/EUROBUS nas empresas do consórcio GIST.

6.1 Estudo dos requisitos do negócio

Como o Módulo QIO está inserido no sistema GIST98/EUROBUS, será analisada a área de planeamento e controlo dos serviços das viaturas e do pessoal tripulante, aliás, a única para a qual o sistema referido tem informação relevante. Uma vez escolhida a área de trabalho é necessário definir uma metodologia para o estudo dos requisitos do negócio. A metodologia que de seguida se descreve é estritamente empírica.

O primeiro passo consistiu em contactos preliminares com as várias empresas do consórcio para, dessa forma, recolher informação. Foram particularmente relevantes o relatório de exploração mensal da STCP [STCP, 1998] e as propostas de indicadores a incluir no Módulo QIO apresentadas pelas empresas STCP, Horários do Funchal, Rodoviária de Lisboa e EVA Transportes, sendo estas duas últimas empresas pertencentes ao Grupo Barraqueiro (anexo A).

A partir da informação recolhida foi possível observar que cada indicador potencial tem no máximo quatro parâmetros. Isto significa que é possível definir com exactidão um indicador respondendo a não mais do que quatro questões:

1. "Qual a designação do indicador?"
2. "A quem (ou a quê) é que o indicador se refere?"
3. "Qual o nível de agregação utilizado?"
4. "Que função de agregação utiliza?"

Por exemplo, para o caso do tempo médio de duração das viagens planeadas por percurso, a designação do indicador é o tempo de duração, refere-se a viagens planeadas, sendo essas viagens agregadas por percurso utilizando a função de cálculo de médias.

Usando esta metodologia foi possível identificar o grupo de indicadores, o grupo de entidades a que os indicadores se referem assim como os níveis e funções de agregação utilizados. Ao reunir estes elementos observou-se que se um indicador pode ser agregado por uma determinada entidade, então também pode ser agregado por entidades que se definam como sendo um conjunto das primeiras entidades referidas. Prosseguindo com o exemplo anteriormente utilizado, se é possível saber o tempo de duração das viagens planeadas por percurso, também é possível saber o tempo de duração das viagens planeadas por linha, dado que uma linha se define como sendo um conjunto de percursos. Fica, assim, estabelecido o conceito de dimensão tal como foi anteriormente definido (subsecção 2.1.6). Usando esta metodologia foi possível construir a primeira versão das dimensões e dos níveis que as constituem (Tabela 6.1). Esses níveis estão ordenados por grau de detalhe. Em primeiro lugar estão os níveis mais gerais, surgindo de seguida os restantes níveis por ordem crescente de detalhe.

Dimensões	Níveis						
Categorias	Categoria	Motorista					
Escalas	Tipo de rede	Estação	Trabalho	Linha	Serviço	Etapa	Dia
Espaço	Tipo de rede	Estação	Megalinha	Linha	Percurso	Segmento	
Grupos	Estação	Grupo	Motorista				
Tempo	Ano	Semestre	Trimestre	Mês	Semana	Tipo de dia	Período
Viaturas	Tipo de rede	Estação	Viatura	Viagem			

Tabela 6.1: 1ª versão das dimensões

Nas tabelas de indicadores, a primeira coluna apresenta as designações desses indicadores. Na linha respectiva indicam-se as dimensões pelas quais os indicadores podem ser agregados. A entidade a que o indicador se refere é apresentada juntamente com a designação desse indicador. A explicitação da entidade é relevante para a equipa de desenvolvimento, sendo só em alguns casos pertinente para os elementos das empresas. De facto, indicar a entidade ‘viagem’ associada ao indicador ‘nº de’ é fundamental para a sua compreensão. No entanto, para a compreensão do indicador ‘distância útil’ é suficiente saber as dimensões pelas quais esse indicador pode ser agregado, assim como a respectiva função de agregação, isto na perspectiva dos elementos da empresa que serviram de interlocutores nesta fase da etapa de análise. Não se encontram representadas as funções de

agregação utilizadas por simplificação. A Tabela 6.2 e a Tabela 6.3 mostram a primeira versão dos indicadores a incluir no Módulo QIO. Por ser a primeira versão, existem várias incorrecções quer na perspectiva do negócio quer na representação descrita. Essa representação foi sendo adaptada e consolidada ao longo das várias versões. A divisão dos indicadores em dois grupos deveu-se à necessidade sentida de diminuir a dimensão do problema. De notar que a única dimensão comum aos dois grupos de indicadores é a dimensão ‘tempo’, mostrando uma relevante diferença estrutural entre estes dois grupos.

	Dimensões						Planeado vs realizado
	Espaço	Tempo	Viaturas	Escalas	Categorias	Grupos	
Extensão	x						
Nº de percursos	x						
Comprimento dos percursos	x						
Nº de linhas	x						
Comprimento das linhas	x						
Nº de viagens	x	x	x	x			x
Nº de turnos		x	x				x
Nº de turnos em simultâneo	x	x	x				x
Distância total	x	x	x	x			x
Distância útil	x	x	x	x			x
Distância em vazio	x	x	x	x			x
Tempo útil	x	x	x	x			x
Tempo em vazio	x	x	x	x			x
Velocidade comercial em linha	x	x					x

Tabela 6.2: 1ª versão dos indicadores de exploração

	Dimensões						Planeado vs realizado
	Espaço	Tempo	Viaturas	Escalas	Categorias	Grupos	
Práticas					x	x	
Nº de motoristas					x	x	
Absentismo					x	x	
Baixas					x	x	
Idade					x		
Tempo na empresa		x			x	x	
Tempo na categoria		x			x	x	
Tempo efectivo diurno		x			x	x	x
Tempo efectivo nocturno		x			x	x	x
Tempo global diurno		x			x	x	x
Tempo global nocturno		x			x	x	x
Tempo extra diurno		x			x	x	x
Tempo extra nocturno		x			x	x	
Tempo teórico		x			x	x	x
Tempo de agente único					x	x	x

Tabela 6.3: 1ª versão dos indicadores sobre motoristas

Nesta 1ª versão, são considerados indicadores de exploração aqueles indicadores que permitem caracterizar o serviço prestado pela empresa tal como é perceptível pelos utentes.

Os indicadores sobre os motoristas incluem tanto os indicadores relevantes para o departamento de recursos humanos como os necessários à equipa responsável pela distribuição do serviço diário dos motoristas. É importante notar que, nas empresas de transportes colectivos, esta tarefa não é, normalmente, realizada por aquele departamento.

Utilizando a primeira versão das dimensões e indicadores a incluir no Módulo QIO, iniciou-se uma série de reuniões com as diferentes empresas do consórcio que foram dando origem a sucessivas versões intermédias das dimensões e indicadores até se alcançar a versão final que se apresenta em Tabela 6.4, Tabela 6.5 e Tabela 6.6.

Comparando a Tabela 6.1 com a Tabela 6.4 observam-se desde logo duas diferenças fundamentais: o maior detalhe dos dados na primeira versão (exemplo: informação por viagem e por segmento) e o maior número de dimensões na versão final. A primeira diferença justifica-se pelo aumento de ‘realismo’ à medida que se aprofundam os conhecimentos técnicos sobre sistemas de Armazenamento de Dados e sobre o próprio negócio, enquanto a segunda é justificada pela abertura de novas perspectivas na análise da informação à medida que aumenta o conhecimento na área de negócios em estudo.

Dimensões	Níveis			
	Ano	Trimestre	Mês	Dia
Datas	Ano			
Datas da rede	Data da rede			
Datas dos horários ao público	Data do horário			
Empresas	Empresa			
Épocas do ano	Época do ano			
Estações de recolha	Estação de recolha			
Grupos de folga	Grupo de folga			
Períodos do dia	Período do dia			
Rede	Linha	Percurso		
Tipos de dia	Tipo de dia			
Tipos de percurso	Tipo de percurso			
Tipos de rede	Tipo de rede			
Tipos de situação	Tipo de situação			
Tipos de viagem	Tipo de viagem			
Tripulação	Categoria	Tripulante		
Viaturas	Turno	Etapa		

Tabela 6.4: Versão final das dimensões

A dimensão empresas é particularmente importante para grupos empresariais constituídos por mais do que uma empresa de transportes colectivos, como é o caso, por exemplo, do Grupo Barraqueiro. Este grupo é constituído, entre outras empresas, pela Rodoviária de Lisboa e pela Transportes a Sul do Tejo, empresas que irão utilizar o sistema GIST98/EUROBUS. Com esta dimensão, os gestores do Grupo Barraqueiro podem

discriminar os indicadores por empresa e efectuar as análises comparativas que acharem pertinentes.

A dimensão ‘tipos de rede’ permite discriminar os indicadores por tipo de rede, como por exemplo, rede de autocarros ou rede de eléctricos.

A dimensão ‘rede’ estabelece uma ordem natural entre linhas e percursos. O conceito ‘megalinha’ é utilizado essencialmente pelos utilizadores do sistema, não sendo usado de forma generalizada pelos restantes elementos da empresa não tendo sido, por isso mesmo, considerado na versão final dos indicadores.

A dimensão ‘datas da rede’ permite identificar as datas em que foram efectuadas alterações à rede comercial.

A dimensão ‘tipos de percurso’ permite distinguir os percursos exteriores dos percursos que pertencem a linhas.

A dimensão ‘estações de recolha’ permite discriminar os indicadores por estação de recolha. Esta dimensão é particularmente importante na análise da informação sobre o serviço das viaturas e dos tripulantes para as empresas que tenham, ao nível do planeamento, o processo de tomada de decisões descentralizado pelas diferentes estações de recolha.

A dimensão ‘viaturas’ compreende dois níveis: turno (ou viatura lógica ou chapa) e etapa. Etapa é uma parte do turno delimitada pela saída da estação de recolha e o regresso a essa ou outra estação de recolha. Nos casos em que um turno regressa à mesma estação de recolha de onde saiu (o que é o mais frequente), é imediata a determinação da estação de recolha à qual o turno pertence. Caso isso não aconteça, o turno é associado à estação de recolha de onde saiu.

As dimensões ‘épocas do ano’ e ‘tipos de dia’ permitem seleccionar a informação, respectivamente, por época do ano e por tipo de dia.

A dimensão ‘tipos de viagem’ permite discriminar as viagens em função do seu tipo. Por exemplo, viagens ao público, viagens em vazio, etc. É de notar que os indicadores ‘distância útil’ e ‘distância em vazio’ referidos na Tabela 6.2 passam a ser referidos, na versão final, como sendo o indicador ‘distância das viagens’ que é agregável, entre outras, pela dimensão ‘tipos de viagem’. Agregando o indicador pelo membro ‘vazio’ da referida

dimensão, obtém-se a distância em vazio; agregando pelo membro ‘ao público’, obtém-se a distância útil.

A dimensão ‘datas dos horários ao público’ permite identificar as datas de novas versões dos horários de viaturas disponibilizadas aos potenciais utentes.

A dimensão ‘datas’ permite seleccionar informação por ano, trimestre, mês ou dia.

A dimensão ‘períodos do dia’ permite discriminar a informação por período do dia.

A dimensão ‘tripulação’ permite discriminar a informação por tripulantes com a possibilidade de os agrupar pelas categorias a que esses tripulantes pertencem.

A dimensão ‘tipos de situação’ aplica-se à informação sobre os tripulantes. Permite filtrar a informação em função de um tipo de situação aplicável aos tripulantes.

A dimensão ‘grupos de folga’ foi introduzida por sugestão da CARRIS tendo em vista a detecção de desequilíbrios entre os grupos de folgas que conduzam a custos adicionais para a empresa, nomeadamente pela existência de um número excessivo de horas extraordinárias. De facto, se o número de tripulantes associados a cada grupo de folga não estiver equilibrado com o serviço a prestar, podem ocorrer situações de excesso de tripulantes nuns dias e falta de tripulantes noutros dias, implicando, neste último caso, o recurso a trabalho extraordinário o que é, naturalmente, mais dispendioso para a empresa.

Na construção da versão final das dimensões houve a preocupação de garantir a conformidade entre todas as dimensões, nomeadamente evitando a existência de diferentes dimensões com níveis comuns. Essa preocupação é bastante importante para assegurar homogeneidade na análise dos diferentes Cubos (ver Tabela 6.5). Foi por este motivo que, por exemplo, foram criadas as dimensões ‘empresas’ e ‘tipos de rede’ quando, uma delas poderia ter sido implementada como o nível de menor detalhe da dimensão ‘rede’. A conformidade entre as várias dimensões é sugerida por [Kimball et al, 1998] como forma de assegurar a integração dos diferentes Armazéns de Dados Departamentais. Este assunto será abordado com mais detalhe na secção 7.3.

Nas tabelas com a versão final dos indicadores equivalentes à Tabela 6.2 e à Tabela 6.3 observa-se que todas as células ficam preenchidas, já que se colocam na mesma tabela os indicadores que se agregam pelo mesmo conjunto de dimensões. Fica, assim, estabelecido o conceito de Cubo tal como foi anteriormente definido (subsecção 2.1.6). Uma vez que todas as células ficam preenchidas e que pela Tabela 6.5 já se sabe quais as dimensões

pelos quais os indicadores de cada Cubo podem ser agregados, optou-se pela utilização de outra forma de representação onde se apresenta a lista de indicadores a disponibilizar em cada Cubo (Tabela 6.6).

Dimensões \ Cubos	Rede comercial	Serviços das viaturas	Serviços das viaturas ao longo do dia	Pessoal tripulante
Datas		x	x	x
Datas da rede	x			
Datas dos horários ao público		x	x	
Empresas	x	x	x	x
Épocas do ano		x	x	
Estações de recolha		x	x	x
Grupos de folga				x
Períodos do dia			x	
Rede	x	x	x	
Tipos de dia		x	x	
Tipos de percurso	x			
Tipos de rede	x	x	x	x
Tipos de situação				x
Tipos de viagem		x	x	
Tripulação				x
Viaturas		x	x	

Tabela 6.5: Relação das dimensões por Cubo

Rede comercial	Serviços das viaturas	Serviços das viaturas ao longo do dia	Pessoal tripulante
Nº de percursos	Nº de viagens	Nº de turnos	Nº de situações planeadas para os tripulantes
Extensão dos percursos	Distância das viagens		Nº de situações realizadas pelos tripulantes
	Tempo de condução nas viagens		Tempo efectivo planeado para os tripulantes
	Tempo de condução nas viagens incluindo paragem nos terminos		Tempo efectivo realizado pelos tripulantes
	Velocidade comercial nas viagens		Tempo extraordinário planeado para os tripulantes
	Velocidade de exploração nas viagens		Tempo extraordinário realizado pelos tripulantes
			Tempo nocturno planeado para os tripulantes
			Tempo nocturno realizado pelos tripulantes

Tabela 6.6: Versão final dos indicadores

Os indicadores referentes à rede comercial são fortemente condicionados por não estarem disponíveis os dados geográficos com informação sobre os percursos na Base de Dados de informação geográfica. Só assim será possível obter a partir do sistema GIST98/EUROBUS, por exemplo, a extensão da rede comercial. De facto, usando os dados da Base de Dados de planeamento, este indicador não pode ser obtido, já que não é

garantida a correspondência entre a rede de planeamento tal como é definida no Módulo Rede e a rede física. Não é possível, por exemplo, diferenciar dois segmentos com sentidos contrários correspondentes a dois troços de estrada paralelos de outros dois segmentos com sentidos contrários mas referentes a um mesmo troço de estrada ou, se existem troços de estrada comuns em diferentes segmentos, etc. Da mesma forma, não é possível obter o indicador ‘comprimento das linhas’ referido na Tabela 6.2. O indicador ‘extensão dos percursos’ agregado pelo nível ‘linha’ da dimensão ‘rede’, usando a função de agregação ‘soma’, dá como resultado a soma da extensão dos percursos por linha, o que é claramente distinto do indicador anteriormente referido.

A informação sobre os serviços das viaturas circunscreve-se ao serviço planeado. O sistema GIST98/EUROBUS não tem informação sobre as viaturas físicas, não sendo assim possível obter os indicadores para o serviço realizado pelas viaturas. Mesmo assim, seria possível saber, por exemplo, o número de quilómetros realizados por dimensão ‘rede’. Essa informação pressupõe que, ao guardar o serviço realizado pelos tripulantes, são também armazenadas as viagens associadas a esse serviço, o que não se verifica.

A possibilidade de agrupar os indicadores tanto pela dimensão ‘datas’ como pela dimensão ‘datas dos horários ao público’ explica-se com os exemplos que se seguem: para saber a distância total planeada para cada um dos diferentes horários ao público que já estiveram em vigor, utiliza-se a dimensão ‘datas dos horários ao público’ para agrupar os dados; para saber a distância total planeada por mês (independentemente do número de versões de horários ao público que estiveram em vigor), utiliza-se a dimensão ‘datas’.

Se os indicadores ‘nº de viagens’ e ‘distância das viagens’ não levantam dúvidas de interpretação, em relação aos restantes indicadores é necessário fazer algumas observações. O tempo de condução nas viagens e o tempo de condução nas viagens com paragem nos terminos diferem porque, no último caso, é contabilizado o tempo que as viaturas estão paradas nas paragens de terminos das viagens. A importância destes dois indicadores reside no facto de o primeiro ser especialmente relevante para a percepção do serviço prestado pela empresa por parte dos utentes, enquanto o segundo permite avaliar o esforço usado (neste caso em termos do recurso tempo) para a prestação desse mesmo serviço. A diferença entre os indicadores ‘velocidade comercial’ e ‘velocidade de exploração’ tem a mesma origem. Assim, enquanto a primeira pode ser interpretada como a velocidade percebida pelo utente, a segunda reflecte a velocidade percebida pela empresa.

O grau de maior detalhe da dimensão ‘datas’ é o dia. No entanto, esta opção impede a obtenção de alguns indicadores da maior relevância para os agentes de decisão. Um desses indicadores é o número de viaturas a utilizar em simultâneo ao longo do dia, pois permite estimar quantas viaturas são necessárias para satisfazer a oferta de viagens. Para contornar esta limitação optou-se por utilizar a dimensão ‘períodos do dia’. Assim, os indicadores sobre os serviços das viaturas ao longo do dia podem ser vistos como uma extensão dos indicadores sobre os serviços das viaturas onde constam os indicadores que têm de ser detalhados por período do dia.

A informação sobre os tripulantes é fortemente condicionada por nem sempre se saber quais as viagens a realizar/realizadas por cada tripulante. Esta limitação impede, por exemplo, a obtenção do tempo de condução planeado/realizado por motorista, o nº de quilómetros planeados/realizados por motorista, etc. Assim, os indicadores disponíveis estão limitados a tempos de serviço planeado/realizado e à identificação da situação planeada/realizada atribuível a cada tripulante em cada dia, como por exemplo, em serviço, de férias, de baixa, etc.

Se os conceitos ‘tempo extraordinário’ e ‘tempo nocturno’ são de compreensão imediata, com o conceito ‘tempo efectivo’ não se passa o mesmo. Assim, tempo efectivo é o tempo em que o tripulante está ao serviço da empresa. O tempo efectivo pode incluir tanto tempo extraordinário como tempo nocturno. Como as empresas não podem atribuir tempo extraordinário aos tripulantes sem o seu consentimento, o processo mais comum é o de definir os serviços diários a realizar pelos tripulantes sem tempo extraordinário, sendo afixados os serviços não cobertos (ou seja, aqueles que não têm tripulantes para os realizar). Os tripulantes podem candidatar-se à realização desses serviços tendo, caso lhes sejam atribuídos esses serviços, mais do que um serviço diário para realizar. Assim, o tempo efectivo pode corresponder à soma dos tempos efectivos de mais do que um serviço diário. O mesmo pode acontecer para o caso dos tempos extraordinário e nocturno.

6.2 Estudo das fontes de dados

Uma vez definidos os indicadores é necessário saber como são obtidos a partir das fontes de dados disponíveis. Como no caso do Módulo QIO as únicas fontes de dados são as Bases de Dados do sistema GIST98/EUROBUS e destas, apenas a Base de Dados de planeamento estava suficientemente desenvolvida durante a fase de concepção e

desenvolvimento do Módulo QIO, as fontes de dados ficam reduzidas a uma ou mais Bases de Dados de planeamento.

A base de trabalho utilizada foi o conjunto de esquemas relacionais da Base de Dados de planeamento criados pelos elementos da equipa de desenvolvimento com recurso à ferramenta *CASE (Computer Aided Software Engineering) ORACLE Designer 2000* (ver anexo B). Uma vez que os dados disponíveis na Base de Dados de planeamento do sistema GIST98/EUROBUS podem ser facilmente agrupados em função dos módulos do sistema onde esses dados são definidos optou-se, numa primeira fase, por identificar a origem (ou fonte) dos dados a incluir no Módulo QIO, pelos módulos do sistema GIST98/EUROBUS onde eles se definem (Tabela 6.7).

Origem / Destino	Novo	Rede	Mega-linhas	Viagens e Viaturas	Administração	Gestão de Pessoal Tripulante	Geração de Escalas	Pós-Escalamento
Datas	x						x	
Datas da rede	x							
Datas dos horários ao público				x				
Empresas	x							
Épocas do ano		x						
Estações de recolha					x			
Grupos de folga					x			
Linhas			x					
Percursos			x					
Períodos do dia	x							
Tipos de dia		x						
Tipos de rede	x							
Tipos de situação					x			
Tipos de viagem				x				
Tripulação					x	x		
Viaturas				x				
Indicadores da rede comercial		x	x					
Indicadores dos serviços das viaturas		x	x	x				
Indicadores dos serviços das viaturas ao longo do dia		x	x	x				
Indicadores do pessoal tripulante						x	x	x

Tabela 6.7: Módulos de origem dos dados para a Base de Dados de análise e controlo

Durante a elaboração da tabela com o esquema da origem dos dados para o Módulo QIO surgiram diversas dificuldades das quais se destacam:

- As datas em que cada horário ao público foi utilizado não estão registadas na Base de Dados de planeamento. Se um horário diz respeito, por exemplo, ao tipo de dia

‘feriados’, é necessário afectar ao dia ‘5 de Outubro’ os horários para esse tipo de dia. Podem existir dias que, não sendo feriados, tenham uma oferta idêntica à de um feriado. Essa situação acontece, tipicamente, caso ocorram ‘pontes’ com a previsão de grande taxa de adesão. Assim, é necessário saber, para cada dia, quais os horários ao público utilizados.

- As datas da rede não ficam registadas na Base de Dados de planeamento. O processo de aquisição de dados deverá ser capaz de detectar alterações efectuadas à rede comercial e de registar a data de ocorrência.
- As empresas e os tipos de rede não constam da Base de Dados de planeamento mas sabe-se que cada Base de Dados de planeamento diz respeito apenas a uma empresa e a um tipo de rede (subsecção 4.2.8). O Módulo QIO terá, necessariamente, informação sobre as várias Bases de Dados de planeamento de onde carrega os dados necessários. A cada uma dessas Bases de Dados deverá estar associada informação da empresa e do tipo de rede respectivos.
- Existem dados necessários ao Módulo QIO que têm de ser introduzidos directamente na Base de Dados de análise e controlo. Esses dados estão assinalados na coluna ‘Novo’ da Tabela 6.7.

Os períodos do dia podem ser obtidos a partir da Base de Dados de planeamento (subsecção 4.2.1). No entanto, uma vez que os períodos tal como são definidos no Módulo Rede podem variar em função da megalinha, e também porque podem existir vários períodos para espaços de tempo demasiado curtos (especialmente nas horas de ponta), esses períodos não se adequam para o fim a que se destina a dimensão ‘períodos do dia’. De facto, esta dimensão tem como objectivo principal saber o número de viaturas ao longo do dia permitindo, em particular, saber o número máximo de viaturas necessárias em cada dia. Assim, é mais adequada a definição de períodos do dia com intervalos de tempo iguais e únicos, ou seja, aplicáveis a todas as megalinhas.

Destaca-se ainda a ausência de dados com origem nos módulos de serviços, de pré-escalamento e de informação ao público. No caso dos dois primeiros módulos referidos, tal facto deve-se ao carácter transitório da informação neles contida. Assim, os conceitos ‘serviço’ e ‘serviço de escala’ são passos preliminares no processo de definição do serviço diário de cada tripulante sendo apenas relevante, na perspectiva da análise da

qualidade do serviço planeado, este último conceito que é definido no Módulo Geração de Escalas. No caso do Módulo de Informação ao Público, o atraso na obtenção de dados geográficos e no consequente desenvolvimento do módulo comprometeu a utilização desses dados no Módulo QIO até ao momento de escrita desta dissertação.

Numa segunda fase construíram-se a Tabela 6.8 e a Tabela 6.9 onde se detalha a origem dos dados. Nestas tabelas são identificadas as tabelas das Bases de Dados fonte necessárias para obtenção dos dados de cada uma das dimensões e de cada um dos indicadores (ver anexo B). É importante clarificar os critérios utilizados durante a fase de preenchimento. No caso das dimensões com um único nível o seu preenchimento não apresenta dificuldades. No entanto, para o caso de dimensões com dois ou mais níveis é necessário identificar a origem dos dados necessários para efectuar a ligação entre os vários níveis. Essa informação fica, por convenção, associada ao nível de maior detalhe. No caso dos indicadores são identificadas todas as tabelas necessárias à sua obtenção para os níveis de maior detalhe das dimensões que lhes são associáveis. Caso os indicadores possam ser calculados a partir de outros indicadores apresentam-se os indicadores necessários a esse cálculo.

Dimensão →Nível	Origem dos dados
Datas	Processo de aquisição de dados e <i>Shifts_schedules; Daily_rosters</i>
Datas da rede →Data da rede	Processo de aquisição de dados
Datas dos horários ao público →Data do horário	Processo de aquisição de dados
Empresas →Empresa	Preenchimento manual
Épocas do ano →Época do ano	<i>Year_seasons</i>
Estações de recolha →Estação de recolha	<i>Depots</i>
Grupos de folga →Grupo de folga	<i>Freeday_groups</i>
Períodos do dia →Período do dia	<i>Periods</i>
Rede →Linha	<i>Lines</i>
Rede →Percurso	<i>Paths e Lines_paths</i>
Tipos de dia →Tipo de dia	<i>Day_types</i>
Tipos de percurso →Tipo de percurso	Preenchimento manual
Tipos de rede →Tipo de rede	Preenchimento manual
Tipos de situação →Tipo de situação	<i>Situation_types</i>
Tipos de viagem →Tipo de viagem	Preenchimento manual
Tripulação →Categoria	<i>Categories</i>
Tripulação →Tripulante	<i>Crew_members</i>
Viaturas →Turno	<i>Shifts_schedules, Shifts e Trips</i>
Viaturas →Etapa	<i>Shifts_schedules, Shifts, Trips, Trips_nodes, Glines_nodes_types e Nodes_types</i>

Tabela 6.8: Tabelas de origem dos dados para as dimensões do Módulo QIO

Dada a pequena dimensão do Módulo QIO e da equipa de desenvolvimento não se sentiu necessidade de detalhar mais a origem dos dados. No entanto, para o caso de sistemas de

maior dimensão e que envolvam uma equipa de desenvolvimento multi-disciplinar, seria aconselhável a elaboração de um documento com a descrição detalhada da forma de cálculo de cada um dos níveis das dimensões e dos indicadores.

Cubo →Indicador	Origem dos dados
Rede comercial →Nº de percursos	<i>Paths e Lines_paths</i>
Rede comercial →Extensão	<i>Paths, Paths_segs e Segments</i>
Serviços das viaturas →Nº de viagens	<i>Trips, Shifts e Shifts_schedules</i>
Serviços das viaturas →Distância	<i>Trips, Shifts, Shifts_schedules, Paths, Paths_segs e Segments</i>
Serviços das viaturas →Tempo de condução	<i>Trips, Shifts e Shifts_schedules</i>
Serviços das viaturas →Tempo de condução incluindo paragem nos terminos	<i>Trips, Shifts, Shifts_schedules, Trips_nodes, Nodes_types e Glines_nodes_types</i>
Serviços das viaturas →Velocidade comercial	Serviços das viaturas →Distância e Serviços das viaturas →Tempo de condução
Serviços das viaturas →Velocidade de exploração	Serviços das viaturas →Distância e Serviços das viaturas →Tempo de condução incluindo paragem nos terminos
Serviços das viaturas ao longo do dia →Nº de turnos	<i>Shifts e Shifts_schedules</i>
Pessoal tripulante →Nº de situações planeadas	<i>Daily_situations e Situation_types</i>
Pessoal tripulante →Nº de situações realizadas	<i>Daily_situations e Situation_types</i>
Pessoal tripulante →Tempo efectivo planeado	<i>Roster_lines, Daily_rosters, Daily_duties e Rostering_duties</i>
Pessoal tripulante →Tempo efectivo realizado	<i>Roster_lines, Daily_rosters, Daily_duties e Rostering_duties</i>
Pessoal tripulante →Tempo extraordinário planeado	<i>Roster_lines, Daily_rosters, Daily_duties e Rostering_duties</i>
Pessoal tripulante →Tempo extraordinário realizado	<i>Roster_lines, Daily_rosters, Daily_duties e Rostering_duties</i>
Pessoal tripulante →Tempo nocturno planeado	<i>Roster_lines, Daily_rosters, Daily_duties e Rostering_duties</i>
Pessoal tripulante →Tempo nocturno realizado	<i>Roster_lines, Daily_rosters, Daily_duties e Rostering_duties</i>

Tabela 6.9: Tabelas de origem dos dados para os indicadores do Módulo QIO

6.3 Desenho da arquitectura do Armazenamento de Dados

Para desenhar a arquitectura do Armazenamento de Dados começa-se por estudar a dimensão da Base de Dados. Em função dos resultados obtidos nesse estudo foram escolhidas as ferramentas a analisar e definidos os critérios de avaliação a adoptar na análise dessas ferramentas. É, de seguida, apresentada a avaliação efectuada e, por fim, definida a arquitectura do Armazenamento de Dados do Módulo QIO.

6.3.1 Estudo da dimensão da Base de Dados

Para o correcto desenho de um sistema de Armazenamento de Dados é importante introduzir dois conceitos novos:

- Janela de Oportunidade: “é o horário, num período de 24 horas, em que existe disponibilidade para a realização de operações de manutenção, nomeadamente operações para a aquisição de dados e para a execução de cópias de segurança”.
- Base de Dados de Grande Dimensão: “é uma Base de Dados em que a Janela de Oportunidade é insuficiente para a realização das operações de manutenção” [Corey et al, 1999].

Para o correcto desenho de um sistema de Armazenamento de Dados é importante determinar se o Armazém de Dados se enquadra, ou não, na definição de Base de Dados de Grande Dimensão. As opções técnicas adoptadas para Bases de Dados de Grandes Dimensões são mais complexas condicionando o desenho da arquitectura do Armazenamento de Dados.

O espaço ocupado pela Base de Dados depende naturalmente do esquema de desenho adoptado. O modelo mais comum designa-se por esquema em estrela e é pressupondo este desenho na construção da Base de Dados que se fizeram os cálculos para estimar o espaço ocupado. Tal opção deve-se ao facto de este modelo de desenho ser aquele que mais espaço consome permitindo, assim, uma estimativa prudente.

O esquema em estrela é constituído por uma tabela de factos para cada Cubo e tantas tabelas de dimensões quanto o número de dimensões associadas a esse Cubo. As chaves primárias das tabelas de dimensões são constituídas por apenas um campo enquanto a chave primária da tabela de factos é constituída por tantos campos quanto o número de dimensões. Na tabela de factos existe ainda um campo por cada indicador do Cubo enquanto em cada tabela de dimensões existem os campos necessários para caracterizar os vários níveis da dimensão (Figura 6.1).

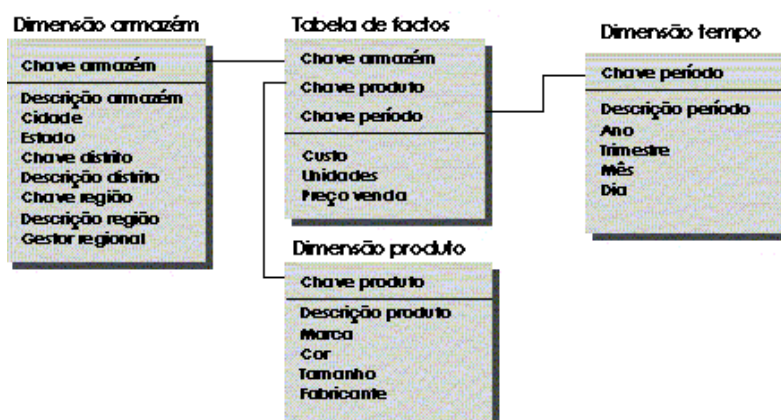


Figura 6.1: Esquema em estrela (traduzido de [Raden, 1995/6])

Para estimar o espaço ocupado pelos dados necessários a cada Cubo, e uma vez que ainda não se sabe quais as ferramentas a utilizar, recorreu-se à seguinte fórmula de cálculo:

$$fmidx \times fmdiv \times (ech \times ndim + eind \times nind) \times nregmin \times fmgra$$

fmidx – factor de multiplicação para estimar o espaço ocupado pelos índices necessários ao bom funcionamento da Base de Dados: 2.

fmdiv – factor de multiplicação para estimar o espaço ocupado por diversos dados tais como o dicionário de dados, dados sobre as dimensões ou espaço alocado que é deixado livre por diferentes motivos: 1,5.

ech – espaço ocupado por cada campo pertencente a uma chave primária: 6 Bytes.

ndim – número de dimensões independentes entre si: assinaladas com **X** na Tabela 6.10.

eind – espaço ocupado por cada indicador: 4 Bytes.

nind – número de indicadores do Cubo: ver Tabela 6.6.

nregmin – nº mínimo de registos necessários (utilizaram-se valores aproximados dos utilizados pela empresa CARRIS): é o produto cartesiano do número de membros dos níveis de maior detalhe das dimensões independentes (assinaladas com **X** na Tabela 6.10).

fmgra – factor de multiplicação referente ao número médio de registos necessários por cada unidade de *nregmin*. Pressupõe que se estabeleça, desde já, a granulosidade dos dados a armazenar. Tendo em conta as observações feitas na subsecção 2.1.5, é necessário encontrar o compromisso entre o espaço ocupado pelos dados e o grau de detalhe pretendido para a informação. Para os Cubos ‘rede comercial’, ‘serviços das viaturas ao longo do dia’ e ‘pessoal tripulante’, a granulosidade dos dados é um reflexo directo do grau de detalhe com que os dados são analisados. No entanto, é importante notar que se perde, desta forma, informação sobre a rede comercial pois o grau de maior detalhe sobre a rede é o percurso quando poderia ser o segmento. Esta opção deve-se ao facto de os segmentos não terem, do ponto de vista do negócio, grande significado dado que a sua definição é efectuada pelos planeadores por razões de ordem estritamente operacional. Quanto ao Cubo sobre os serviços das viaturas, é importante guardar a informação das viagens, mesmo que não se obtenha indicadores por viagem, pois tendo esses dados é possível a qualquer momento calcular novos indicadores. O nº médio de

viagens por cada etapa de uma viatura estima-se em cerca de 15. Será esse o valor de *fmgra* para o Cubo serviços das viaturas. Para os restantes Cubos esse valor será de 1.

Dimensões	Nº	Rede comercial	Serviços das viaturas	Serviços das viaturas ao longo do dia	Tripulantes
Datas	365		X	X	X
Datas da rede	10	X			
Datas dos horários ao público	10		x	x	
Empresas	1	X	X	X	X
Épocas do ano	3		x	x	
Estações de recolha	6		x	x	x
Grupos de folga	10				x
Períodos do dia	96			X	
Rede	350	X	x	x	
Tipos de dia	3		x	x	
Tipos de percurso	2	x			
Tipos de rede	1	X	X	X	X
Tipos de situação	100				x
Tipos de viagem	2		x	x	
Tripulação	2500				X
Viaturas	2000		X	X	
Nº de registos:		3.500	10.950.000	70.080.000	912.500
Espaço ocupado (Mb):		0,32	1.503,75	6.817,02	146,20

Tabela 6.10: Estimativa empírica do espaço a ocupar pela Base de Dados de análise e controlo por ano

Antes de analisar os resultados obtidos convém notar que os valores dados a *fmidx* e *fmdiv* foram obtidos com base em pequenas amostras e/ou experiência empírica. É, por isso mesmo, fundamental estudar com mais cuidado valores médios para esses factores de multiplicação de modo a poder-se utilizar a fórmula apresentada num âmbito mais alargado. No entanto, para o caso concreto do sistema desenvolvido, estes valores vieram a mostrar-se adequados.

Analisando os resultados apresentados na Tabela 6.10 pode-se observar que cerca de 80% do espaço total ocupado é utilizado para a obtenção do indicador ‘nº de turnos’. Utilizando um espaço temporal de 3 anos (número de anos considerado suficiente pelas empresas de transportes para a realização de análises temporais) o espaço ocupado pelos restantes indicadores é de cerca de 4,8 *Giga Bytes*. Uma vez que o indicador ‘nº de turnos’ pode ser obtido por cálculo desde que se saibam as horas de início e de fim de cada etapa das viaturas, os cerca de 20 *Giga Bytes* por ano necessários para a obtenção desse indicador pode ser reduzido de forma muito significativa. De facto, o valor de *nind* do Cubo ‘serviços das viaturas’ aumenta 2 (correspondentes ao armazenamento dos indicadores ‘hora de início’ e ‘hora de fim’ das viagens) implicando um aumento no espaço ocupado

por ano pelo Cubo ‘serviços das viaturas’ de 1,5 para 1,7 *Giga Bytes* deixando de ser necessário o Cubo ‘serviços das viaturas ao longo do dia’. Ou seja, em 3 anos, o espaço ocupado total passaria a ser de 5,6 *Giga Bytes* em vez dos anteriores 24,8 *Giga Bytes*.

Utilizando, por exemplo, uma *DAT (Digital Audio Tape)* de 4mm, o tempo necessário para fazer uma cópia de segurança integral será de pouco mais de uma hora e meia. O tempo em que as fontes de dados estão disponíveis variam, naturalmente, de empresa para empresa. Utilizando como exemplo o caso em que esse tempo é mais reduzido, as fontes de dados estão disponíveis da 1 hora da manhã até às 6 horas da manhã, em concorrência com outras aplicações que efectuam exclusivamente operações de leitura sobre essas fontes de dados, e das 6 às 9 horas da manhã, com disponibilidade total por parte dessas mesmas fontes de dados. Uma vez que a realização de cópias de segurança da Base de Dados de análise e controlo não depende da disponibilidade das fontes de dados, pode ser efectuada, por exemplo, das 22 às 24 horas. Quanto às operações de aquisição de dados têm, para o exemplo em análise, cerca de 8 horas. Uma vez que o volume diário de dados a carregar dificilmente ultrapassará os 6 *Mega Bytes* (valor obtido por recurso à folha de cálculo apresentada na Tabela 6.10 utilizando o espaço temporal de um dia) e dada a relativa simplicidade de processamento, o tempo disponível parece ser claramente suficiente. Ou seja, a Base de Dados a construir não se enquadra na definição de Base de Dados de Grande Dimensão.

6.3.2 Selecção das ferramentas *OLAP* a analisar

Na selecção das ferramentas a analisar procurou-se encontrar um conjunto de ferramentas com diferentes abordagens técnicas. Analisando a Tabela 6.11 pode-se ver como se posicionam as ferramentas de maior expressão no mercado, segundo os critérios definidos na subsecção 2.2.3.

Destas ferramentas seleccionaram-se as seguintes:

- Uma ferramenta de referência de entre as que utilizam arquitectura *ROLAP*: *DSS Agent and Server* da *MicroStrategy, Inc.*;
- Uma ferramenta de referência de entre as que utilizam arquitectura *MOLAP*: *Express Server* da *Oracle Corporation*;
- Uma ferramenta com soluções técnicas inovadoras: *TM1* da *Applix, Inc.*;

- Uma ferramenta de baixo custo e com várias opções quanto à forma de armazenamento: *SQL Server 7/OLAP Services* da *Microsoft Corporation*.

Opções de processamento multi-dimensional	Forma de armazenar os dados		
	Base de Dados relacional	Base de Dados multi-dimensional	Ficheiros localizados no cliente
SQL em passos múltiplos	Cartesis Carat MicroStrategy		
Servidor OLAP	Sagent EUREKA: Strategy Longview Khalix Informix MetaCube Speedware Media/MR Microsoft OLAP Services Oracle Express (ROLAP mode) Pilot Analysis Server IBM DB2 OLAP Server Seagate Holos (ROLAP mode) Applix TM1 WhiteLight	SAS CFO Vision Comshare Decision Hyperion Essbase Gentia Speedware Media/M Microsoft OLAP Services Oracle Express Pilot Analysis Server PowerPlay Enterprise Server Seagate Holos Applix TM1	
Cliente OLAP	Oracle Discoverer Informix MetaCube	Comshare FDC Dimensional Insight Hyperion Enterprise Hyperion Pillar PwC CLIME	Brio Enterprise BusinessObjects Cognos PowerPlay Personal Express iTM1 Perspectives

Tabela 6.11⁷: Matriz de caracterização das ferramentas OLAP (traduzido de [Pendse & Creeth, 2000a])

6.3.3 Avaliação das ferramentas OLAP

Os critérios de avaliação utilizados [Demarest, 1995], [Hackney, 1997], [Pendse & Creeth, 1997] podem ser divididos em dois grupos: os que permitem avaliar as ferramentas na perspectiva do utilizador final e aqueles que permitem a avaliação na perspectiva da equipa de desenvolvimento. No primeiro grupo constam como critérios a fiabilidade do Sistema de Gestão de Bases de Dados, a variedade e a qualidade das aplicações de utilização amigável e produção de relatórios e o preço. No segundo grupo constam a capacidade, a simplicidade estrutural da Base de Dados, a flexibilidade estrutural da Base de Dados e a funcionalidade de cálculo, a funcionalidade de acesso a dados e o suporte técnico e a documentação disponíveis.

⁷ A matriz apresentada é de Agosto de 2000 apesar de a caracterização das ferramentas OLAP ter sido efectuada em finais de 1998. No entanto, as diferenças existentes não são significativas dado que as

Para avaliar as ferramentas é necessário, antes de mais, enquadrar a construção do Módulo QIO numa perspectiva de futuro. Este tema será abordado mais em detalhe na secção 7.2; no entanto, deverá ficar desde já claro que o Módulo QIO não pretende ser um ‘embrião’ de um Armazém de Dados, sendo o seu crescimento condicionado pela evolução do sistema GIST98/EUROBUS. Desta forma, não é crítica a capacidade da ferramenta mas, em contrapartida, é crítico o preço da solução escolhida, dada a necessidade de tornar essa solução viável para todas as empresas do consórcio GIST. Houve, ainda, a preocupação de garantir que a soma dos pesos para cada um dos dois grupos atrás referidos é igual. Assim, atendendo à Tabela 6.12, a soma dos pesos para cada um dos grupos é de 230.

CrITÉRIOS de avaliação	Peso	DSS Agent and Server	Data Mart Suite	TM1	SQL Server 7
Capacidade	10	9	9	6	4
Simplicidade estrutural da Base de Dados	50	3	5	8	9
Flexibilidade estrutural da Base de Dados e funcionalidade de cálculo	50	6	8,5	7	7
Funcionalidade de acesso a dados	70	8	10	4	7
Suporte técnico e documentação	50	7	9	5	7
Fiabilidade do Sistema de Gestão de Bases de Dados	70	n.a. (8)	7	8	4
Aplicações de utilização amigável e produção de relatórios	70	7	8	5,5	8
Preço	90	6	2	9	10
PONTUAÇÃO TOTAL	4600	3040	3145	3095	3420
Classificação		66,09%	68,37%	67,28%	74,35%

n.a. - não aplicável por não ter um sistema gestor de Bases de Dados próprio

Pontuações de 1 a 10

Tabela 6.12: Avaliação das ferramentas *OLAP*

A valorização de cada um dos critérios foi efectuada recorrendo a [Pendse & Creeth, 1997], a apresentações efectuadas das ferramentas em análise e, pontualmente, a testes (Tabela 6.12). Dos valores apresentados chama-se a atenção para o valor atribuído à fiabilidade do Sistema de Gestão de Bases de Dados do *DSS Agent and Server*. Apesar de esta ferramenta não incluir um Sistema de Gestão de Bases de Dados utilizando para o efeito outros Sistemas de Gestão de Bases de Dados, atribuiu-se o valor de 8 correspondente à avaliação do *ORACLE8 Enterprise Edition* por ser este o Sistema de Gestão de Bases de Dados utilizado pela Base de Dados de planeamento e seria, naturalmente, a opção a utilizar para o Módulo QIO caso se escolhesse o *DSS Agent and Server*. Os valores da capacidade e da fiabilidade do Sistema de Gestão de Bases de Dados do *SQL Server 7* foram baixos por não existirem dados suficientes para uma melhor avaliação desta ferramenta dada a sua juventude no momento da avaliação. Os valores atribuídos ao preço seguiu o seguinte critério: definiram-se intervalos de preço de 750.000\$00 por ordem ascendente e atribuíram-se valores por ordem descendente de 10 até

ferramentas então disponíveis mantém-se no mercado, em alguns casos com novas versões e/ou diferentes

A construção de um sistema de Armazenamento de Dados no âmbito do sistema GIST98/EUROBUS

1. Assim, o valor máximo avaliável seria de 7.250.000\$00 o que, no contexto em que o Módulo QIO se insere é já um valor excessivo.

A ferramenta *OLAP* seleccionada⁸ é, então, o *SQL Server 7*. As aplicações de utilização amigável e de produção de relatórios são deixadas ao critério das empresas do consórcio GIST. No entanto, para efeito de demonstrações, foram utilizados o *Microsoft Excel 2000* e/ou o *Knosys ProClarity 2.0*, que se utilizarão também nesta dissertação para melhor ilustrar o ambiente de trabalho que os utilizadores finais terão ao seu dispor.

6.3.4 Definição da arquitectura geral do sistema

Tendo em conta as ferramentas seleccionadas, a arquitectura geral do sistema de Armazenamento de Dados para o caso do Módulo QIO será tal como se apresenta na Figura 6.2.

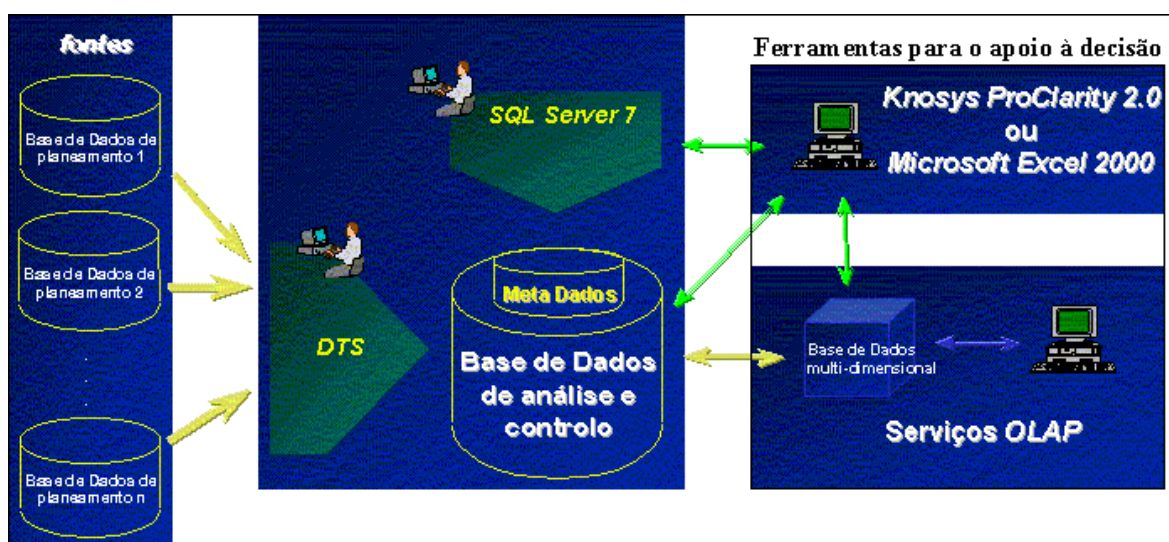


Figura 6.2: Arquitectura do Módulo QIO (adaptado de [Ranito, 1999])

Os serviços de transformação de dados (*DTS – Data Transformation Services*) e os serviços *OLAP* fazem parte integrante do *SQL Server 7*. Os *DTS* (subsecção 6.5.3) permitem implementar as tarefas de migração de dados, nomeadamente o carregamento dos dados, para a Base de Dados relacional que servirá de repositório aos dados necessários para os serviços *OLAP*. Os serviços *OLAP* (subsecção 6.6.1) armazenam as

designações, mas sem alterações face aos critérios de análise utilizados.

⁸ A avaliação, à data de escrita desta dissertação, das novas versões das ferramentas *OLAP* seleccionadas iria reforçar de forma clara a opção tomada [Pendse & Creeth, 2000b]. A classificação em percentagem atribuível ao *SQL Server 2000*, utilizando os mesmos critérios, é de cerca de 90% enquanto a classificação para os restantes produtos não sofreria variações superiores a 2% em valor. Esta diferença justifica-se pela juventude dos serviços *OLAP* em finais de 1998 de que resultou uma sub-avaliação no estudo então efectuado.

agregações que processam na Base de Dados relacional referida (caso se opte pela arquitectura *ROLAP*) ou numa Base de Dados multi-dimensional (caso se opte pela arquitectura *MOLAP* ou *HOLAP*). O utilizador final terá como *software* o *Knosys ProClarity 2.0* ou o *Microsoft Excel 2000* (subsecção 6.6.2) que acedem aos serviços *OLAP*.

O sistema operativo do servidor é o *Microsoft Windows NT Server Enterprise Edition 4.0* ou versão posterior com *SP4* ou versão posterior, enquanto que o sistema operativo dos clientes é o *Windows NT Workstation 4.0* ou versão posterior com *SP4* ou versão posterior, ou ainda, o *Microsoft Windows 95/98* [Microsoft Corporation, 1998]. Não são aqui considerados sistemas operativos para tecnologias não compatíveis com a tecnologia *Intel*.

As configurações do servidor e dos clientes serão apresentadas na fase de integração da solução na empresa cliente, altura em que se poderá estimar com mais rigor a dimensão da Base de Dados.

6.4 Instalação da infra-estrutura de Armazenamento de Dados

A instalação do *SQL Server 7* [Microsoft Corporation, 1998] é bastante intuitiva não tendo sido detectados problemas significativos. Chama-se apenas a atenção para o facto de o processo de instalação ser diferente para o servidor e para os clientes.

6.5 Desenho e construção da Base de Dados e do processo de aquisição dos dados

O desenho da Base de Dados relacional foi efectuado utilizando como referência o modelo dimensional apresentado em [Kimball et al, 1998]. Assim, começa-se por apresentar de forma muito sucinta o essencial do modelo dimensional, de seguida mostra-se a sua aplicação ao caso em estudo e, finalmente, descreve-se o processo de aquisição de dados.

6.5.1 O modelo dimensional

O modelo dimensional utiliza o esquema em estrela (subsecção 6.3.1) como referência no desenho da Base de Dados. No entanto, encontram-se tipificadas situações em que devem ser feitas adaptações ao esquema em estrela:

- Quando uma dimensão tem vários níveis e o número de membros de cada um desses níveis é muito desigual, adopta-se o esquema em cristal de gelo. Por exemplo, o caso de uma empresa com várias unidades fabris onde se utiliza uma dimensão em que o

nível de maior detalhe é a informação do funcionário e o nível imediatamente superior é a unidade fabril onde esse funcionário trabalha. O esquema em cristal de gelo caracteriza-se por colocar em diferentes tabelas cada um dos níveis de uma mesma dimensão (Figura 6.3). O esquema em cristal de gelo tem, em comparação com o esquema em estrela, a vantagem de poupar espaço de armazenamento em situações como a descrita anteriormente. A desvantagem consiste na utilização de mais tabelas, o que pode fazer aumentar os tempos de resposta na obtenção dos indicadores e/ou aumentar o tempo de processamento dos Cubos.

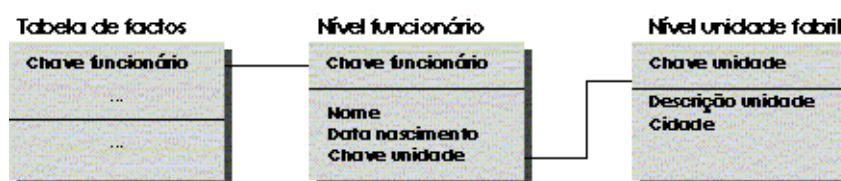


Figura 6.3: Esquema em cristal de gelo

- Quando uma dimensão se altera ao longo do tempo, embora com pouca frequência, podem-se utilizar diferentes opções:
 - Tipo 1: escrever por cima do registo da dimensão perdendo, dessa forma, o valor anterior. Esta opção pressupõe que não é necessário guardar os diferentes valores que uma dimensão toma ao longo do tempo.
 - Tipo 2: criar um novo registo de dimensão com novo código associado, referente a uma nova data. Esta opção deve ser utilizada caso seja relevante guardar os diferentes valores que uma dimensão vai tomando ao longo do tempo.
 - Tipo 3: criar um novo campo para registar o valor imediatamente anterior. Para exemplificar a utilidade desta opção veja-se uma situação em que o regime fiscal é alterado sendo dada ao contribuinte a possibilidade de escolher qual o regime pretendido: o anterior ou o novo. É neste tipo de situação, em que ocorre uma mudança ‘suave’, que se deve utilizar esta opção.
- Quando uma dimensão se altera com muita frequência ao longo do tempo, utiliza-se a opção de tipo 2.

6.5.2 Desenho e construção da Base de Dados [Microsoft Corporation, 1998]

No desenho do esquema relacional para o Cubo ‘rede comercial’ (Figura 6.4) efectuaram-se as seguintes alterações relativamente ao esquema em estrela:

- As chaves estrangeiras referentes às dimensões ‘empresas’ e ‘tipos de rede’ não se encontram na tabela de factos (*network_fact*) mas sim na tabela referente à dimensão ‘rede’ (*network*) pois esta dimensão não é independente das dimensões ‘empresas’ e ‘tipos de rede’, economizando-se desta forma espaço em disco;
- O campo *pathtype_id* não faz parte da chave primária da tabela *network_fact* pois o tipo de um percurso é único numa dada data.

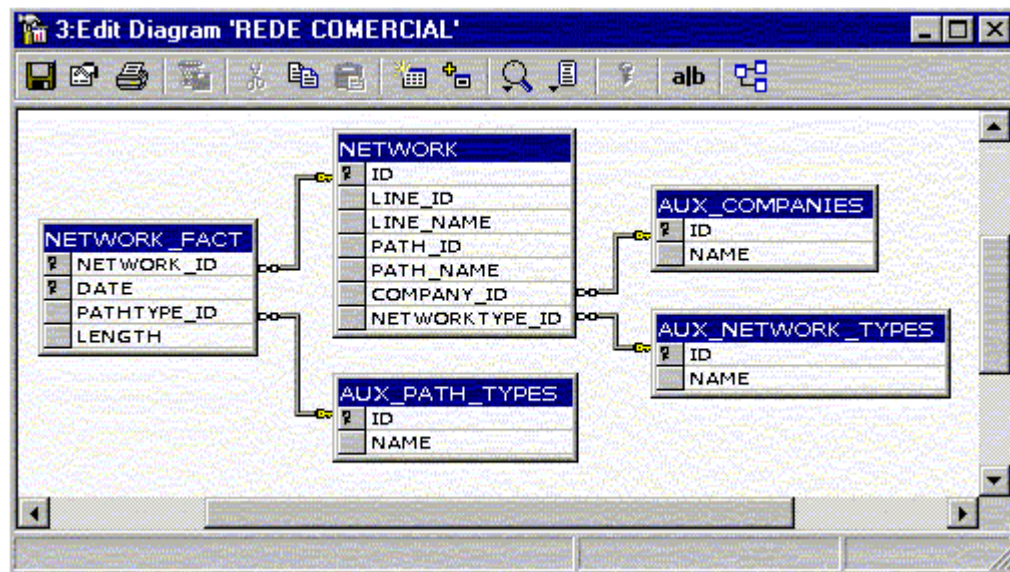


Figura 6.4: Esquema relacional para o Cubo ‘rede comercial’

As alterações feitas ao esquema em estrela para o caso do esquema relacional referente ao Cubo ‘serviços das viaturas’ (Figura 6.5) foram:

- A não colocação do campo *date* na tabela *trips* para evitar redundância de informação;
- Uma vez que a tabela de factos é a tabela *trips*, e dado que sabendo a viagem se sabem univocamente todos os valores para as dimensões que lhe estão associadas (com a excepção já referida da dimensão data), a chave primária da tabela *trips* é constituída por apenas um campo (*id*) que é o identificador da viagem;
- As tabelas *shifts_stretches*, *depots*, *day_types* e *year_seasons* estão previstas no primeiro caso de excepção descrito na apresentação do modelo dimensional (subsecção 6.5.1).

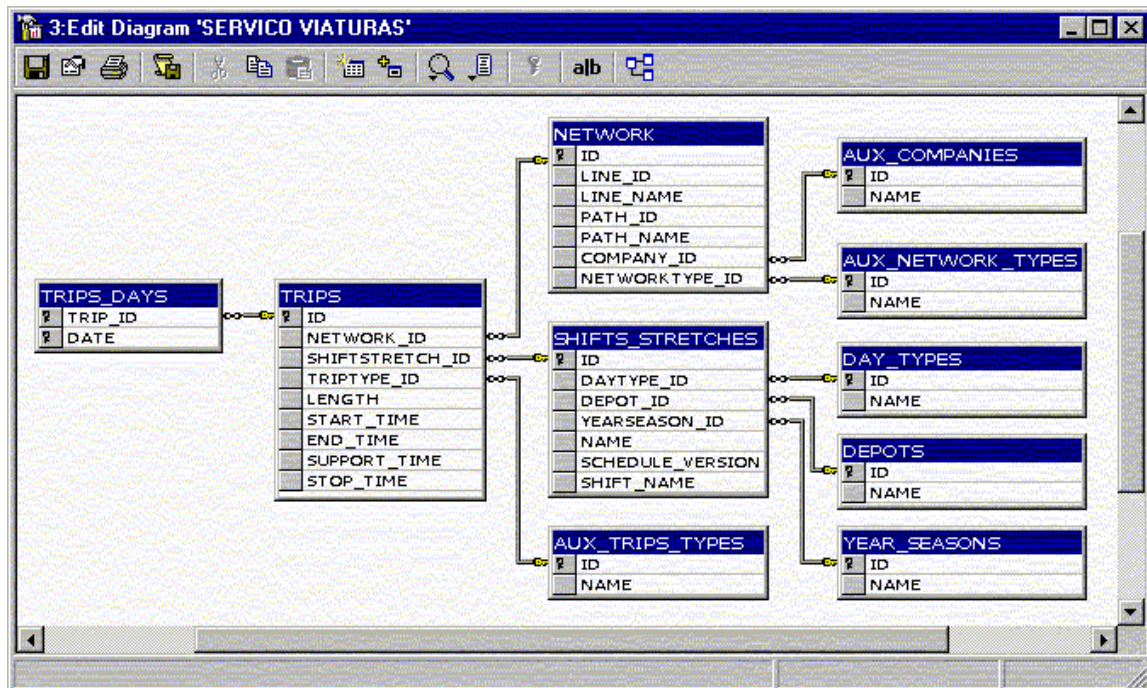


Figura 6.5: Esquema relacional para o Cubo 'serviços das viaturas'

As alterações mais relevantes efectuadas no desenho do esquema referente ao Cubo 'pessoal tripulante' (Figura 6.6) em relação ao esquema em estrela foram:

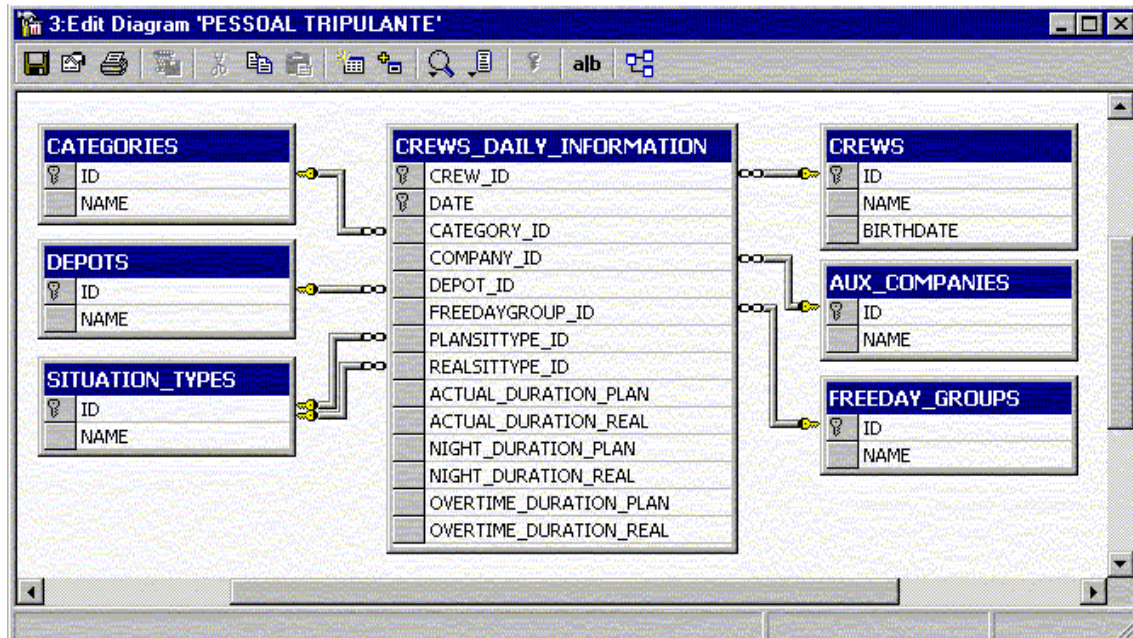


Figura 6.6: Esquema relacional para o Cubo 'pessoal tripulante'

- Os campos *plansittype_id*, *realstitype_id*, *category_id*, *company_id*, *depot_id*, *freedaygroup_id* e *networktype_id* não fazem parte da chave primária da tabela *crews_daily_information* (a tabela de factos) pois todos esses valores são únicos para cada dia de um tripulante;

- A implementação da dimensão ‘tripulação’ tem a particularidade de o nível ‘tripulante’ ser invariável ao longo do tempo enquanto o nível ‘categoria’ pode variar ao longo do tempo. Assim, optou-se por colocar em tabelas diferentes os dados referentes a cada um dos níveis. Os dados sobre os tripulantes encontram-se na tabela *crews* e os dados sobre as categorias encontram-se na tabela *categories*.

6.5.3 O processo de aquisição de dados

O processo de aquisição de dados foi desenvolvido utilizando os *DTS* [Microsoft Corporation, 1998]. Antes de iniciar a descrição da construção do processo de aquisição de dados começa-se por descrever as tabelas criadas para o efeito, organizadas em cinco tipos distintos identificáveis pelas primeiras letras dos seus nomes:

- *Sys_*: tabelas necessárias à gestão do processo de aquisição de dados (Figura 6.7);

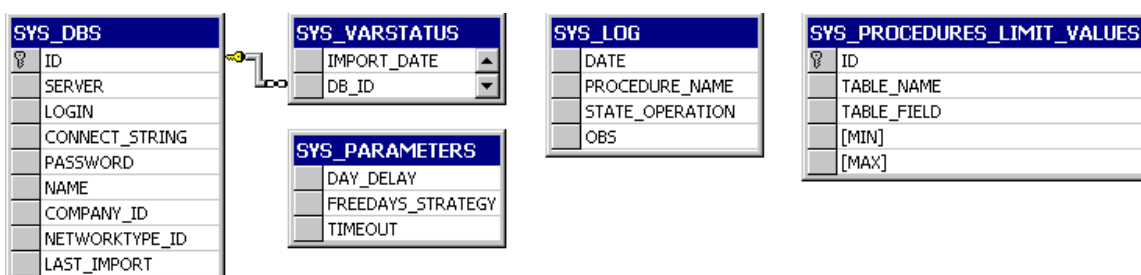


Figura 6.7: Esquema relacional para as tabelas *sys_*

- *Temp_*: tabelas utilizadas para guardar valores temporários durante o processo de aquisição de dados;
- *Map_*: tabelas utilizadas para fazer a correspondência entre as entidades das Bases de Dados fonte e a Base de Dados de análise e controlo (na Figura 6.8 apresenta-se a tabela referente à entidade ‘percurso’, sendo as tabelas referentes às restantes entidades em tudo idênticas);

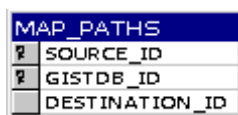


Figura 6.8: Exemplo de uma tabela *map_*

- *Aux_*: tabelas auxiliares que contêm dados necessários ao bom funcionamento do Módulo QIO e que não podem ser obtidas a partir das Bases de Dados fonte sendo, por isso mesmo, preenchidas manualmente (Figura 6.9);

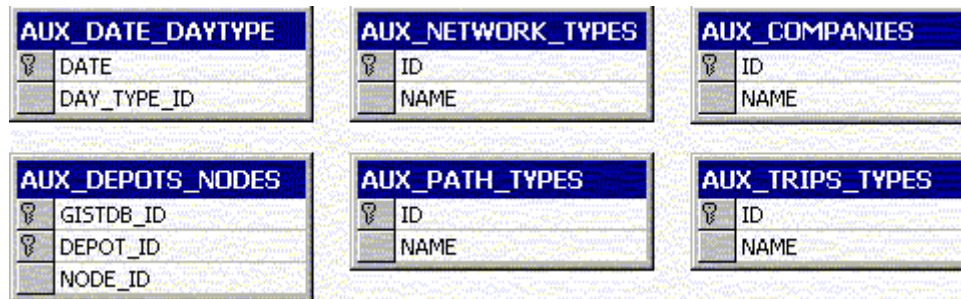


Figura 6.9: Esquema relacional para as tabelas *aux_*

- *Tempqio_*: tabelas utilizadas nas Bases de Dados fonte para auxiliar o processo de aquisição de dados (Figura 6.10). O esquema relacional apresentado na Figura 6.10 foi realizado com recurso à ferramenta *CASE ORACLE Designer 2000*.

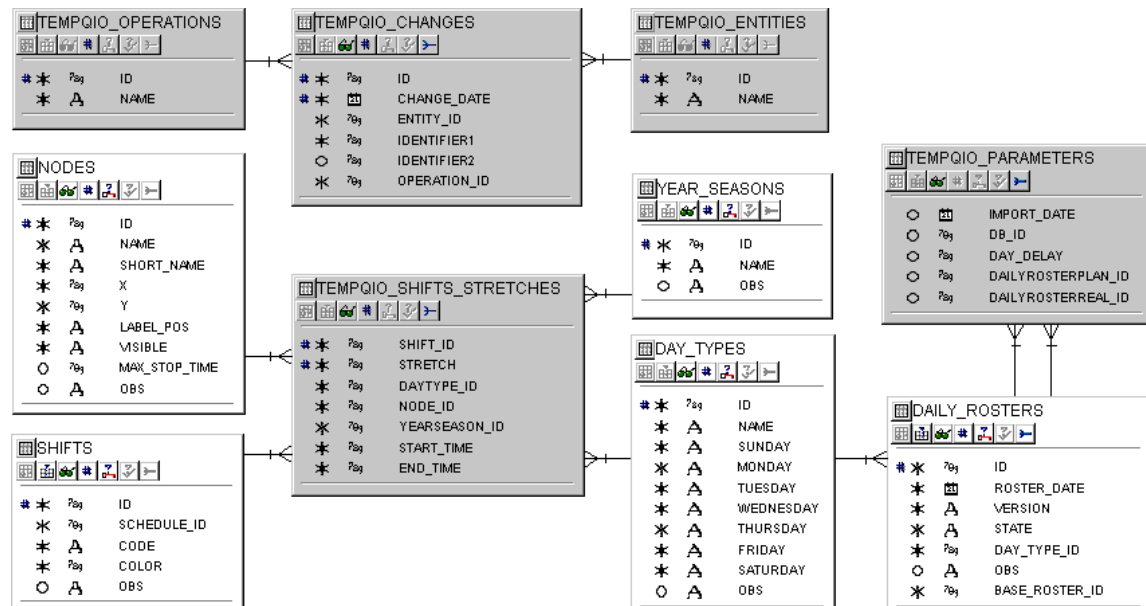


Figura 6.10: Esquema relacional para as tabelas *tempqio_*

Com excepção das tabelas *aux_*, as restantes tabelas apresentadas armazenam Meta Dados (ver subsecção 2.2.2). O papel de cada um dos tipos de tabelas descritos fica mais claro na descrição que a seguir se faz pela ordem estabelecida na subsecção 2.2.1.

Purificação dos dados

Os problemas mais relevantes que se detectaram relativamente à qualidade dos dados fonte, tendo por base dados de três meses da empresa CARRIS, foram:

- A existência de valores inválidos devido, na maior parte dos casos, a erros de introdução de dados;

- A dificuldade em detectar situações em que o mesmo registo passa, num dado momento, a referir-se a uma outra entidade. Exemplificando, se são alterados os dados de um tripulante, essa alteração pode corresponder a uma mera modificação de alguns dados desse tripulante ou pode corresponder à remoção da informação desse tripulante e à sua substituição pela informação de um outro tripulante. Em ambos os casos, mantém-se o mesmo valor da chave primária desse registo.

Para evitar a introdução de valores inválidos, criou-se a tabela *sys_procedures_limit_values* onde estão definidos os valores mínimo e máximo permitidos para cada campo para o qual seja relevante a validação do valor. Os tempos de trabalho do pessoal tripulante foram validados desta forma.

Para evitar que um mesmo registo se passe a referir a uma entidade diferente, foram criadas janelas de aviso no sistema GIST98/EUROBUS onde se informa o utilizador dos erros que podem ocorrer caso utilize a operação de alteração para implicitamente eliminar uma entidade e criar outra. Este aviso existe para as seguintes entidades: tipos de dia, épocas do ano, estações de recolha, percursos, linhas, tipos de viagem, horários de viaturas (Figura 6.11), turnos, tripulantes, categorias, grupos de folga e tipos de situação.

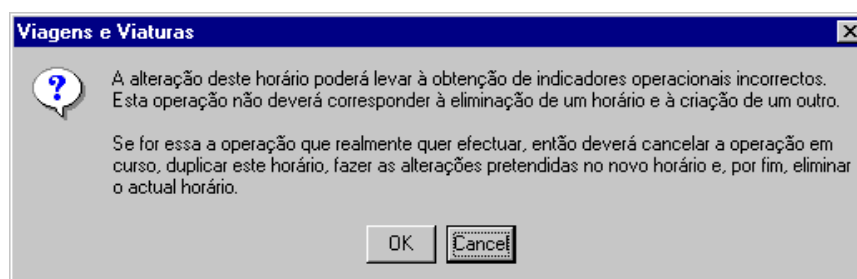


Figura 6.11: Exemplo de uma janela de aviso

Para as entidades que sofrem alterações ao longo do tempo, embora com pouca frequência, é necessário registar a data e o tipo de alterações sofridas para, dessa forma, facilitar a extracção dos dados. Para tal, criou-se nas Bases de Dados fonte a tabela *tempqio_changes* onde é inserida informação dos registos que sofreram alterações. A detecção da ocorrência de alterações e a inserção da informação correspondente nesta tabela é feita por recurso a comandos de activação automática (*triggers*). Usando como exemplo os horários de viaturas (subsecção 4.2.3), o código que a seguir se apresenta (em linguagem *PL/SQL*) garante que, sempre que uma versão é activada ou, estando activa, é desactivada, essa informação fica registada na tabela *tempqio_changes*. É suficiente registar alterações ao campo *active* por não ser possível efectuar alterações sobre versões activas. Ou seja, caso

se alterasse, por exemplo, uma viagem, essa alteração só podia ser efectuada se o respectivo horário de viaturas estivesse desactivado. Nesse caso, quando se tornasse a activar a versão alterada, ficariam registados na tabela *tempqio_changes* a identificação da entidade, a identificação do horário de viaturas, a data de alteração e o tipo de alteração verificada: activação ou desactivação.

```
CREATE OR REPLACE TRIGGER shifts_schedules_changes
BEFORE INSERT OR UPDATE OF active ON shifts_schedules
FOR EACH ROW
WHEN (old.active<>new.active)
DECLARE
    id_sequence NUMBER;
    id_entity NUMBER;
    id_operation INTEGER;
BEGIN
    SELECT id INTO id_entity
    FROM tempqio_entities
    WHERE name = 'SHIFTS_SCHEDULES';
    IF (:new.active='Y') THEN
        SELECT s_tempqio_changes.nextval INTO id_sequence FROM dual;
        SELECT id INTO id_operation FROM tempqio_operations WHERE name = 'ACTIVE';
        INSERT INTO tempqio_changes (id , change_date, entity_id , operation_id, identifier1)
        VALUES (id_sequence , sysdate, id_entity , id_operation , :new.id );
    END IF;
    IF (updating AND (:new.active='N')) THEN
        SELECT s_tempqio_changes.nextval INTO id_sequence FROM dual;
        SELECT id INTO id_operation FROM tempqio_operations WHERE name = 'INACTIVE';
        INSERT INTO tempqio_changes (id , change_date, entity_id , operation_id, identifier1)
        VALUES (id_sequence , sysdate, id_entity , id_operation , :new.id );
    END IF;
END;
```

Para o caso das entidades para as quais não existe o conceito de versão activa como, por exemplo, as linhas, são registadas todas as operações de transacção na tabela *tempqio_changes* como se pode ver no código que se transcreve de seguida.

```
CREATE OR REPLACE TRIGGER lines_changes
BEFORE INSERT OR UPDATE OR DELETE ON lines
FOR EACH ROW
DECLARE
    id_sequence NUMBER;
    id_entity NUMBER;
    id_operation INTEGER;
BEGIN
    SELECT s_tempqio_changes.nextval INTO id_sequence FROM dual;
    SELECT id INTO id_entity FROM tempqio_entities WHERE name = 'LINES';
    IF (inserting) THEN
        SELECT id INTO id_operation FROM tempqio_operations WHERE name = 'NEW';
        INSERT INTO tempqio_changes (id , change_date, entity_id , operation_id, identifier1 )
        VALUES (id_sequence , sysdate , id_entity , id_operation , :new.id );
    END IF;
    IF (updating) THEN
        SELECT id INTO id_operation FROM tempqio_operations WHERE name = 'UPDATE';
```

```

        INSERT INTO tempqio_changes (id , change_date , entity_id , operation_id , identifier1 )
        VALUES (id_sequence , sysdate , id_entity , id_operation , :new.id );
    END IF;
    IF (deleting) THEN
        SELECT id INTO id_operation FROM tempqio_operations WHERE name ='DELETE';
        INSERT INTO tempqio_changes (id , change_date , entity_id , operation_id , identifier1 )
        VALUES (id_sequence , sysdate , id_entity , id_operation , :new.id );
    END IF;
END;
```

Migração dos dados

Na tabela *sys_dbs* encontra-se a informação básica sobre cada uma das Bases de Dados fonte. O campo *last_import* permite saber a que data se referem os últimos dados carregados com sucesso a partir dessa Base de Dados enquanto os campos *company_id* e *networktype_id* permitem associar os dados à empresa e tipo de rede a que essa Base de Dados fonte diz respeito (subsecção 4.2.8). Os restantes campos desta tabela são necessários ao estabelecimento da ligação entre o *SQL Server 7* e a respectiva Base de Dados fonte. Assim, para cada linha da tabela *sys_dbs*, é efectuada a ligação e iniciado o processo de migração dos dados referente ao dia a seguir à data de *last_import* para todos os dias até à data actual menos um período previamente estabelecido. Este período, que se designa por período de maturação dos dados, tem por objectivo garantir que a informação está já suficientemente estável quando é migrada para a Base de Dados de análise e controlo. É na tabela *sys_parameters* que é guardado esse valor (*day_delay*). O campo *timeout* permite estabelecer um tempo máximo para o processo de aquisição de dados. É, assim, possível garantir que não se excede o maior período de tempo contínuo da Janela de Oportunidade (subsecção 6.3.1). Como é óbvio esta tabela tem um só registo.

A tabela *sys_varstatus* tem apenas um registo de cada vez sendo constituída por dois campos: *import_date* e *db_id*. O primeiro campo indica qual a data a que se referem os dados que estão a ser carregados (*last_import* + 1) enquanto o segundo campo indica qual a Base de Dados fonte com que se está a trabalhar. Ou seja, esta tabela foi criada com o objectivo de guardar as variáveis globais do processo de aquisição de dados.

Durante o carregamento dos dados são preenchidas as tabelas designadas por *map_<nome da entidade>*. Estas tabelas são fundamentais na detecção e correcção de erros. São colocados nestas tabelas a identificação da Base de Dados fonte de cada registo, o código do registo na fonte e o código do registo correspondente na Base de Dados de análise e controlo. Os códigos na Base de Dados de análise e controlo são independentes dos

códigos correspondentes nas Bases de Dados fonte por razões de segurança, já que, se isso não acontecesse, os dados da Base de Dados de análise e controlo passariam a estar dependentes de estratégias adoptadas para as Bases de Dados fonte, tais como a reutilização de códigos. É feita a correspondência entre os códigos das Bases de Dados fonte e da Base de Dados de análise e controlo para as seguintes entidades: tipos de dia, épocas do ano, estações de recolha, percursos, linhas, tipos de viagem, turnos, tripulantes, categorias, grupos de folga e tipos de situação.

Também de grande importância na detecção e correcção de erros é a tabela *sys_log*. É possível a partir desta tabela saber quais os procedimentos que se iniciaram e terminaram, assim como os avisos e/ou as falhas que possam eventualmente ter ocorrido. Neste caso é colocada no campo *obs* informação útil para a identificação dessas ocorrências.

Caso o carregamento seja bem sucedido o campo *last_import* da tabela *sys_dbs* é actualizado.

O carregamento das tabelas *aux_* tem de ser feito manualmente. A tabela *aux_depots_nodes* é necessária para saber qual é o nó correspondente a cada uma das estações de recolha definidas nas Bases de Dados de planeamento. A tabela *aux_date_daytype* permite indicar para cada dia qual o tipo de dia associado, por exemplo, o tipo de dia associado ao dia 4 de Novembro de 2000 é o Sábado enquanto o tipo de dia associado ao dia 25 de Dezembro de todos os anos é o Domingo ou feriado. Está previsto que esta informação venha a existir num futuro próximo no sistema GIST98/EUROBUS passando, nessa altura, a ser carregada a partir das Bases de Dados de planeamento e não manualmente. As tabelas *aux_networks_types* e *aux_companies* permitem guardar, respectivamente, os vários tipos de rede com que a empresa opera e a identificação da empresa (caso mais comum) ou das várias empresas (para casos de grupos empresariais constituídos por mais do que uma empresa). A tabela *aux_path_types* tem sempre dois registos. Os dois valores para o campo *name* são ‘exterior’ e ‘em linha’ permitindo, assim, identificar os percursos exteriores e os percursos pertencentes a linhas. O tipo dos percursos está implícito na Base de Dados de planeamento (subsecção 4.2.2). Da mesma forma, a tabela *aux_trips_types* tem apenas dois registos cujos valores do campo *name* são ‘ao público’ e ‘em vazio’ (subsecção 4.2.3). O tipo de cada viagem obtém-se a partir do campo *empty* da tabela *trips* da Base de Dados de planeamento.

A automatização do processo de aquisição de dados é feita recorrendo a uma funcionalidade dos *DTS* que permite definir a hora de início de um processo e a sua periodicidade assim como o período de tempo para o qual a calendarização efectuada é válida (Figura 6.12).

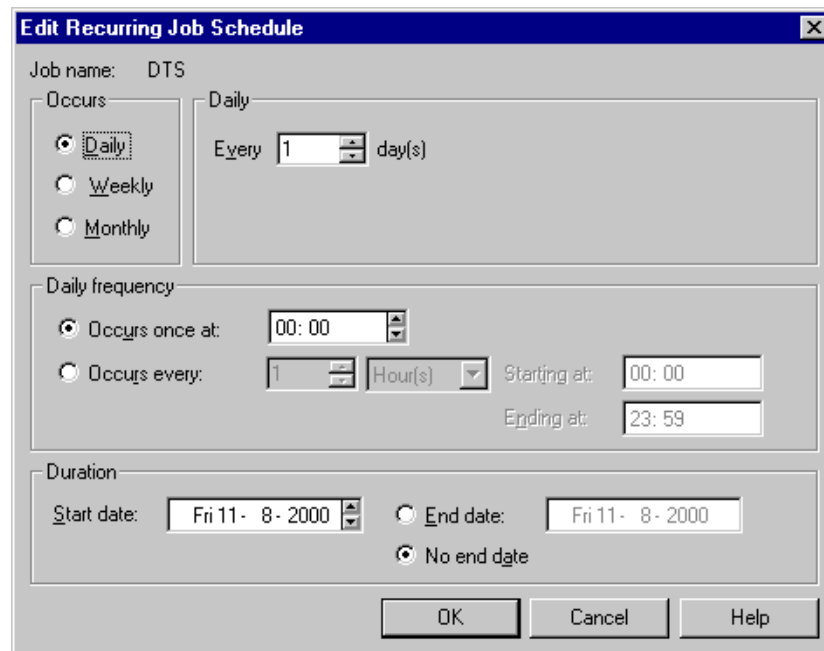


Figura 6.12: Janela para a automatização de um processo

Tendo em atenção a legenda apresentada na Figura 6.13 assim como o esquema de migração dos dados apresentado na Figura 6.14, pode-se desde logo verificar que os *DTS* têm a possibilidade de definir precedências na execução de código existindo duas linguagens que se podem utilizar: *SQL* ou *Visual Basic ActiveX Script*. No entanto, caso se utilize a linguagem *SQL*, só é garantida a precedência do primeiro comando⁹. Para contornar esta limitação pode-se executar um procedimento recorrendo à linguagem *Transact-SQL* garantindo assim a precedência para vários comando *SQL*. A chamada desse procedimento é realizada através do comando *execute*.

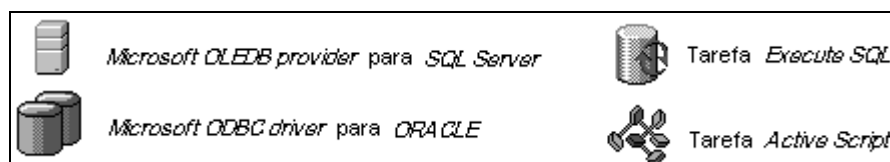


Figura 6.13: Legenda do esquema de migração de dados

⁹ Esta falha foi corrigida no *SQL Server 2000* (a evolução do *SQL Server 7*).

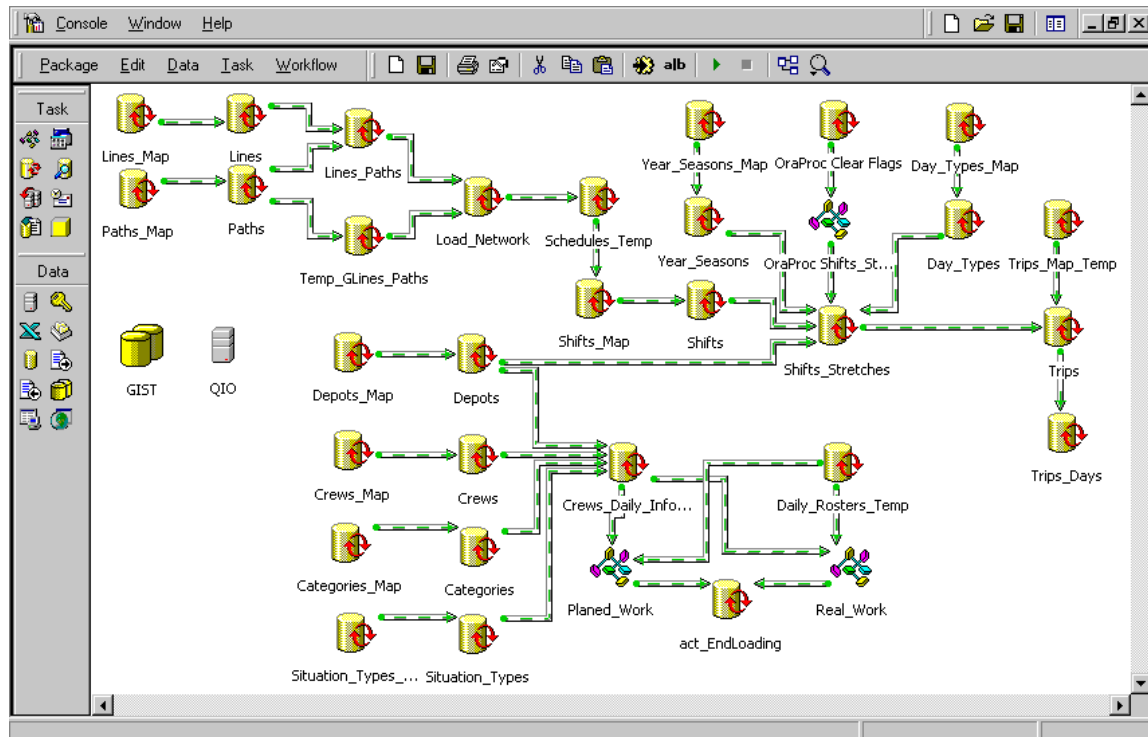


Figura 6.14: Esquema de migração dos dados

Recorre-se à linguagem *Visual Basic ActiveX Script* quando se pretende efectuar alterações sobre Bases de Dados assentes em outras tecnologias, ou seja, no caso do Módulo QIO quando se pretende efectuar alterações nas Bases de Dados fonte. Essa necessidade foi sentida para contornar uma limitação detectada no *SQL Server 7* e que é a dificuldade em executar uma consulta sobre uma Base de Dados remota com passagem de parâmetros. Por exemplo, caso se pretenda seleccionar os tempos de trabalho dos tripulantes para um dia específico, não se consegue passar a data pretendida no comando *SELECT* utilizado para implementar essa consulta. Assim, foi criada nas Bases de Dados fonte a tabela *tempqio_parameters*, que tem os campos necessários ao armazenamento dos valores que se pretende utilizar como parâmetros.

Para facilitar a extracção dos dados foi criada a tabela *daily_informations* nas Bases de Dados fonte (ver anexo B) de forma a simplificar a leitura dos dados referentes ao Módulo Geração de Escalas. Esta tabela contém informação redundante sendo preenchida no módulo referido.

O campo *freedays_strategy* da tabela *sys_parameters* permite identificar a estratégia utilizada para a atribuição dos dias de folga aos tripulantes (subsecção 4.2.4). Este parâmetro é necessário porque o sistema GIST98/EUROBUS suporta duas estratégias

diferentes de atribuição dos dias de folga aos tripulantes, sendo definida a estratégia utilizada pela empresa na configuração do sistema. A leitura dos dados referentes a essa informação depende deste parâmetro.

Falta apenas referir algumas particularidades da migração dos dados referentes aos Cubos ‘rede comercial’ e ‘serviços das viaturas’. Em relação ao Cubo ‘rede comercial’ destaca-se o seguinte:

- Para se registarem alterações efectuadas à rede comercial podem-se utilizar duas estratégias: ou se carregam os dados diariamente ou apenas no caso de haver alterações aos percursos. Optou-se por esta segunda via por economia de espaço e porque não existia qualquer vantagem em ter essa informação numa base diária. A opção utilizada pressupõe o carregamento de toda a informação sobre a rede comercial sempre que haja uma alteração, mesmo que pequena. Quando isso acontece é criada uma nova data de rede.
- Os percursos exteriores (subsecção 4.2.2) são associados a uma linha que existe por defeito e que se designa por linha em vazio.

No caso do Cubo ‘serviços das viaturas’ é relevante guardar o planeamento diário para poder calcular, por exemplo, a distância prevista a percorrer pelas viaturas por mês. No entanto, uma vez que o planeamento do serviço que previsivelmente as viaturas vão percorrer mensalmente se repete para o mesmo tipo de dia e época do ano, essa informação só é carregada para a Base de Dados de análise e controlo quando há alguma alteração, à semelhança do que acontece para o caso da rede comercial. Também neste caso é criada uma nova data dos horários ao público. Assim, apenas a informação da tabela *trips_days* é carregada diariamente.

Para auxiliar a tarefa de transformação dos dados recorre-se às tabelas *temp_* que, como o nome indica, são utilizadas para guardar valores temporários. Assim, estas tabelas só contêm valores durante o processo de aquisição de dados sendo utilizadas na transformação dos dados referentes a qualquer um dos três Cubos. A tabela *tempqio_shifts_stretches* tem uma função semelhante mas para auxiliar os cálculos efectuados nas Bases de Dados fonte, em particular na identificação das etapas dos turnos.

6.6 Instalação e configuração das ferramentas para o apoio à decisão

Uma vez que a instalação dos serviços *OLAP* é feita durante a instalação do *SQL Server 7*, e que as instalações do *Microsoft Excel 2000* e do *Knosys ProClarity 2.0* não apresentam quaisquer dificuldades, passa-se à descrição da configuração dos serviços *OLAP*, do *Microsoft Excel 2000* e do *Knosys ProClarity 2.0*.

6.6.1 A configuração da ferramenta *OLAP*: os serviços *OLAP* do *SQL Server 7*

Os serviços *OLAP* [Microsoft Corporation, 1998] são de utilização muito intuitiva, desde que, naturalmente, o desenho da Base de Dados seja adequado para a construção das dimensões e dos Cubos pretendidos.

A construção de uma dimensão pode ser classificada em dois tipos: esquema em estrela, quando a dimensão utiliza uma só tabela e esquema em cristal de gelo, quando a dimensão utiliza várias tabelas. No primeiro tipo há a considerar um caso especial que é a dimensão tempo. De facto, um só campo do tipo *datetime* ou *smalldatetime* (tipos de dados do *SQL Server 7*) permite caracterizar todos os níveis de uma dimensão tempo. Na construção da dimensão ‘datas’ utilizou-se esta opção permitindo, assim, definir os níveis ‘ano’, ‘trimestre’, ‘mês’ e ‘dia’ utilizando apenas um campo (Figura 6.15). No entanto, uma vez que as datas para o Cubo ‘serviços das viaturas’ são obtidas da tabela *trips_days* enquanto que as datas para o Cubo ‘pessoal tripulante’ se obtêm a partir da tabela *crews_daily_information*, foi necessário construir duas dimensões ‘datas’.

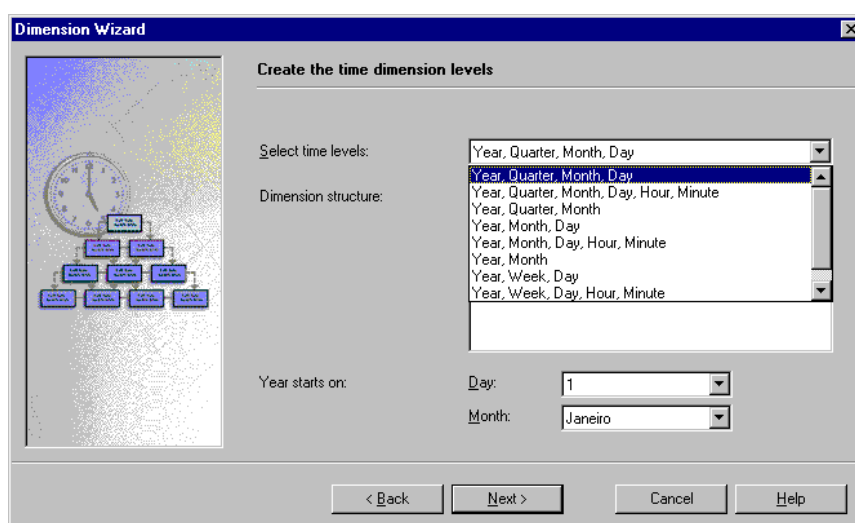


Figura 6.15: Selecção dos níveis de uma dimensão tempo

No caso das dimensões que utilizam várias tabelas (dimensões ‘tripulação’ e ‘viaturas’) é necessário reconstruir as relações entre as tabelas utilizadas já que os serviços *OLAP* não vão buscar essa informação à Base de Dados (Figura 6.16).

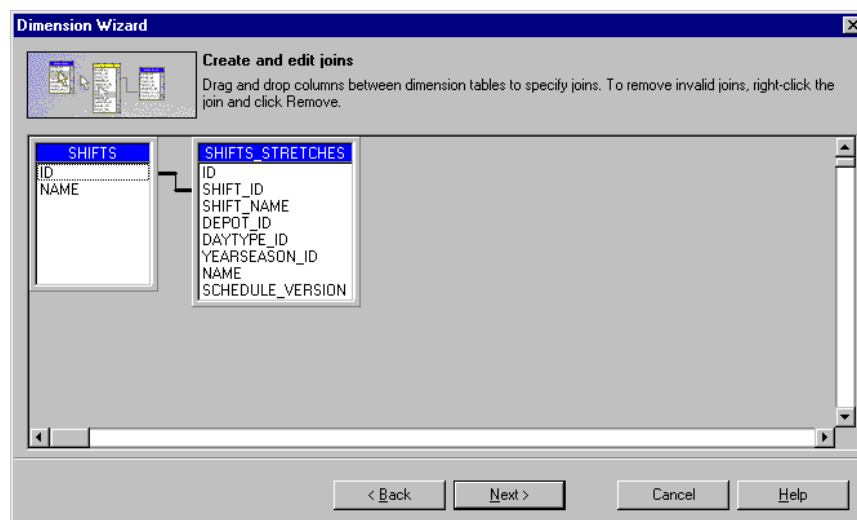


Figura 6.16: Esquema relacional usado na definição de uma dimensão com várias tabelas

Todas as restantes dimensões são de um único nível e, por consequência, utilizam uma só tabela. Apenas uma chamada de atenção para as dimensões ‘datas da rede’ e ‘datas dos horários ao público’ que, apesar de serem definidas com base em campos do tipo *smalldatetime*, não são tratadas como dimensões tempo. De facto, não faz sentido agregar os indicadores que lhes estão associados por outra unidade de tempo para além do dia, sendo suficiente um só nível para caracterizar estas dimensões.

A construção dos Cubos pode ser feita recorrendo a um editor próprio para a execução dessa tarefa ou a um assistente que utiliza (à semelhança do que acontece na definição das dimensões) uma sequência de janelas e que segue, sensivelmente, os seguintes passos:

- Selecção da tabela de factos: um Cubo só pode ter uma tabela de factos. É possível contornar esta limitação através da utilização do conceito de Cubo virtual. No entanto, na construção do Módulo QIO não se revelou necessária a utilização deste conceito não sendo, por isso mesmo, pertinente abordá-lo com mais detalhe.
- Selecção dos campos para a obtenção dos indicadores: pressupõe que um indicador é obtido através de um único campo da tabela de factos utilizando uma função de agregação (por defeito é a função soma, podendo ser alterada recorrendo ao editor de Cubos). Caso se pretenda construir um indicador através de uma função de cálculo mais complexa é necessário recorrer ao editor de Cubos e utilizar uma janela própria

para a construção de indicadores calculados. Utilizou-se esta possibilidade na definição, entre outros, do indicador velocidade de exploração (Figura 6.17).



Figura 6.17: Janela para a construção de indicadores calculados

- Seleção das dimensões: as dimensões podem ser definidas como exclusivas de um determinado Cubo ou partilháveis pelos vários Cubos. É do conjunto das dimensões associáveis ao Cubo em construção que se seleccionam as dimensões efectivamente utilizadas para agregar os indicadores desse Cubo.
- Escolha da forma de armazenamento: das quatro formas de armazenamento dos dados referidas na subsecção 2.2.3 os serviços *OLAP* disponibilizam três dessas formas (Figura 6.18), nomeadamente, a Base de Dados relacional (arquitectura *ROLAP*), a Base de Dados multi-dimensional (arquitectura *MOLAP*) e a Base de Dados híbrida (arquitectura *HOLAP*). A escolha da forma de armazenamento a utilizar pelo Módulo QIO seria, naturalmente, uma opção a tomar durante a etapa de implementação do método de desenvolvimento utilizado para a construção do Módulo QIO já que a quantidade de dados condiciona claramente a opção a tomar.

agregações; e processamento, necessário caso se altere a estrutura do Cubo. O Módulo QIO deverá utilizar o processamento incremental quando estiver em utilização nas empresas.

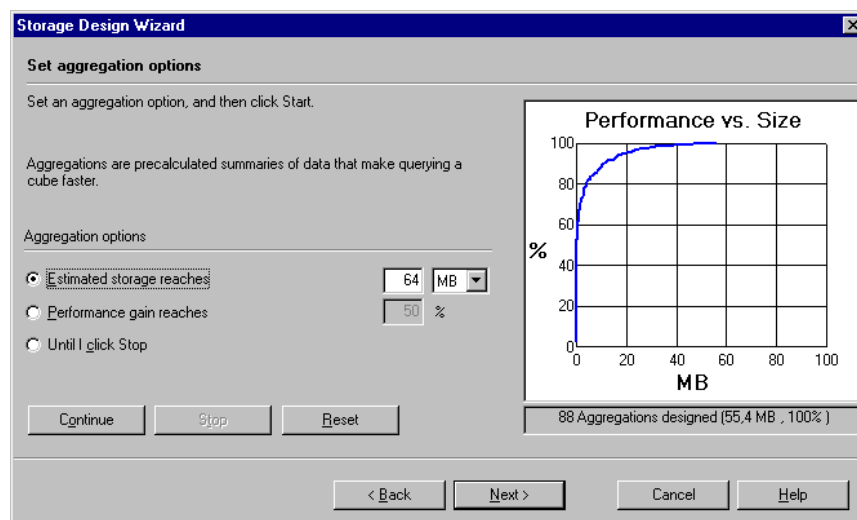


Figura 6.19: Janela para a definição das opções de agregação

6.6.2 Exemplos de ferramentas para o utilizador final

As duas ferramentas utilizadas têm características bem distintas. O *Microsoft Excel 2000* tem como grande vantagem o facto de estar muito difundida e, como tal, ter um custo de aprendizagem baixo por parte dos potenciais utilizadores. Para além disso é bastante flexível na criação de relatórios. O *Knosys ProClarity 2.0* é, fundamentalmente, uma ferramenta de análise sendo bastante versátil quanto às opções que apresenta para o efeito.

Microsoft Excel 2000

Para obter qualquer matriz a partir dos dados dos Cubos, é necessário preencher a janela apresentada na Figura 6.20.

As figuras que se seguem mostram como se pode obter o gráfico com o número de viaturas a utilizar ao longo do dia. Em primeiro lugar é definida a matriz com os valores necessários (Figura 6.20). Com os valores obtidos (Figura 6.21) e ainda com o intervalo de tempo pretendido obtém-se o número de turnos para cada intervalo de tempo usando para tal uma função em linguagem *Visual Basic for Applications*. Foi com base nesses valores que se construiu o gráfico (Figura 6.22) usando as funcionalidades do *Microsoft Excel 2000* próprias para o efeito.

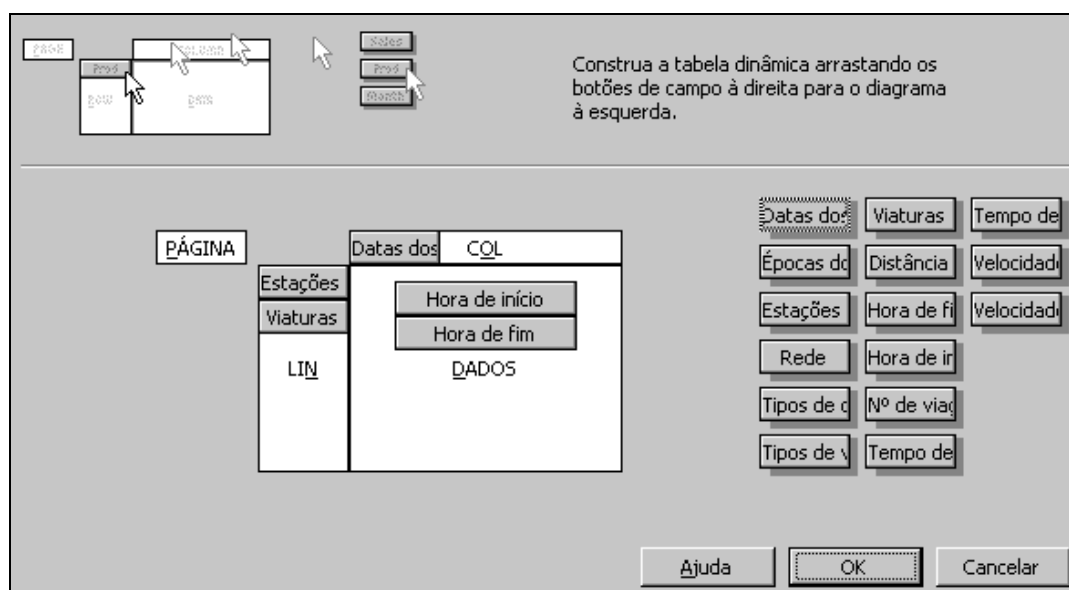


Figura 6.20: Microsoft Excel 2000 - janela para a definição de uma matriz

Estação de recolha	Turno	Etapa	Dados	Data do horário
Cabo Ruivo	15:17-21:29	15:17-21:29	Hora de início	917
			Hora de fim	1289
	15:45-20:18	15:45-20:18	Hora de início	945
			Hora de fim	1218
	15:48-20:35	15:48-20:35	Hora de início	948
			Hora de fim	1235
	15:49-22:32	15:49-22:32	Hora de início	949
			Hora de fim	1352
	15:53-24:42	15:53-24:42	Hora de início	953
			Hora de fim	1482
	15:57-25:53	15:57-25:53	Hora de início	977
			Hora de fim	1538
	16:01-22:22	16:01-22:22	Hora de início	989
			Hora de fim	1322
	16:02-20:36	16:02-20:36	Hora de início	992
			Hora de fim	1216
	16:02-21:05	16:02-21:05	Hora de início	962
			Hora de fim	1265

Figura 6.21: Microsoft Excel 2000 - matriz com dados do Cubo 'serviços das viaturas'

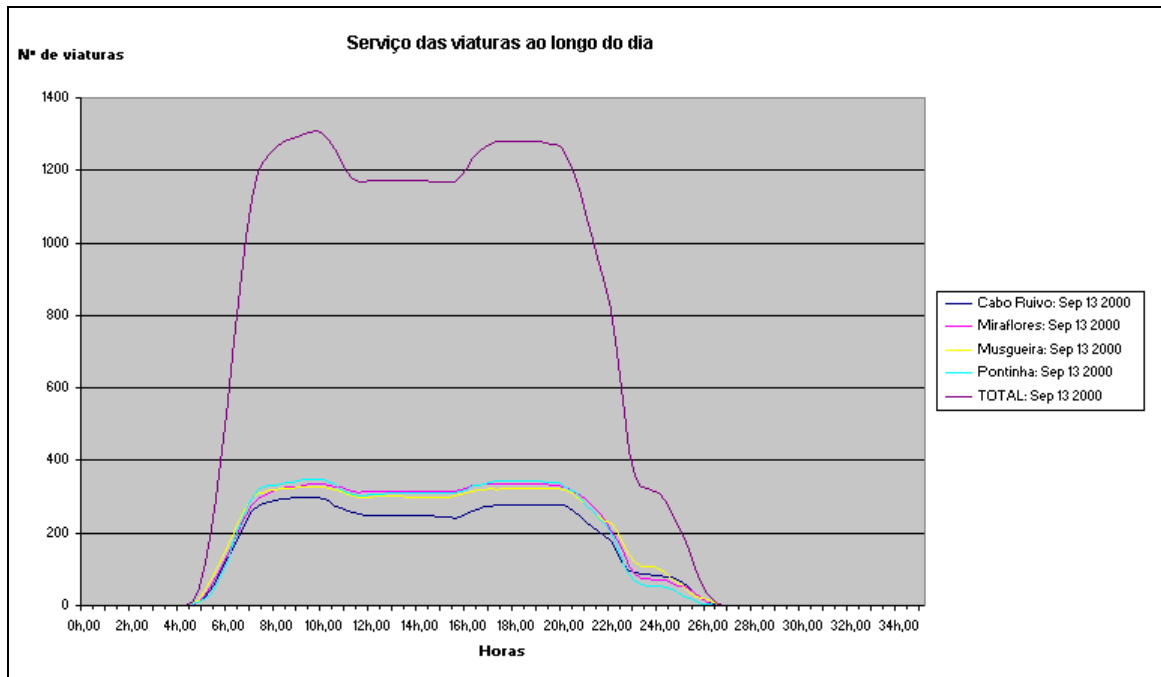


Figura 6.22: *Microsoft Excel 2000* - gráfico com o serviço das viaturas ao longo do dia

Knosys ProClarity 2.0

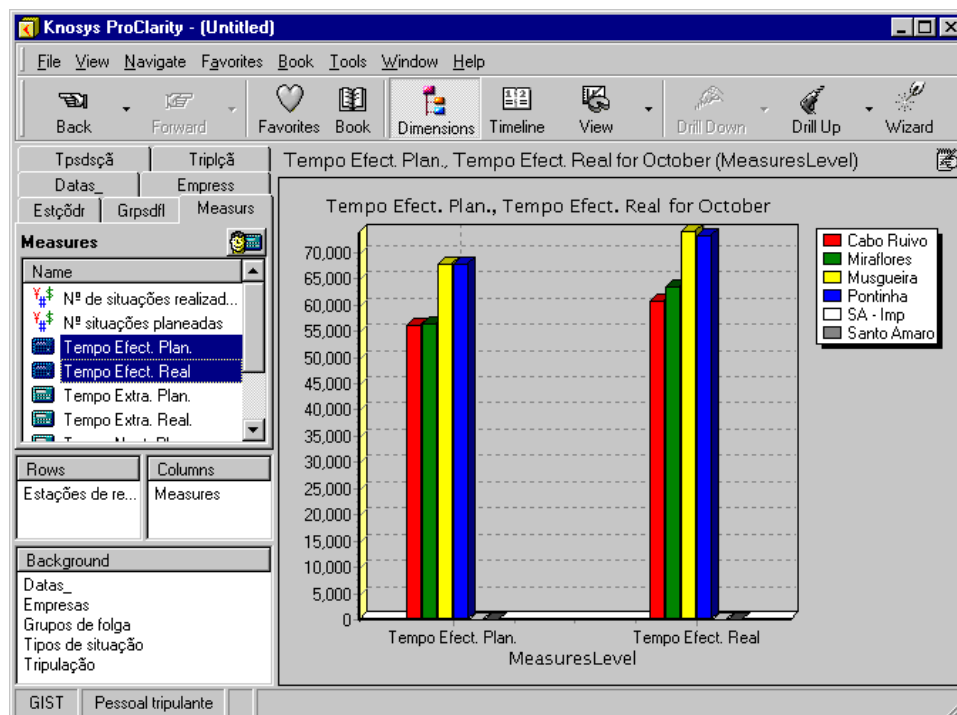
As figuras que se apresentam de seguida pretendem ilustrar alguns dos instrumentos de análise disponíveis no *Knosys ProClarity 2.0*.

Na Figura 6.23 mostra-se a visualização em formato de matriz, podendo-se ver do lado esquerdo os indicadores (*measures*) seleccionados. O indicador ‘Tempo Efect. Real. / Plan.’ foi definido usando funcionalidades próprias do *Knosys ProClarity 2.0* para o efeito. Os dados apresentados dizem respeito ao mês de Outubro.

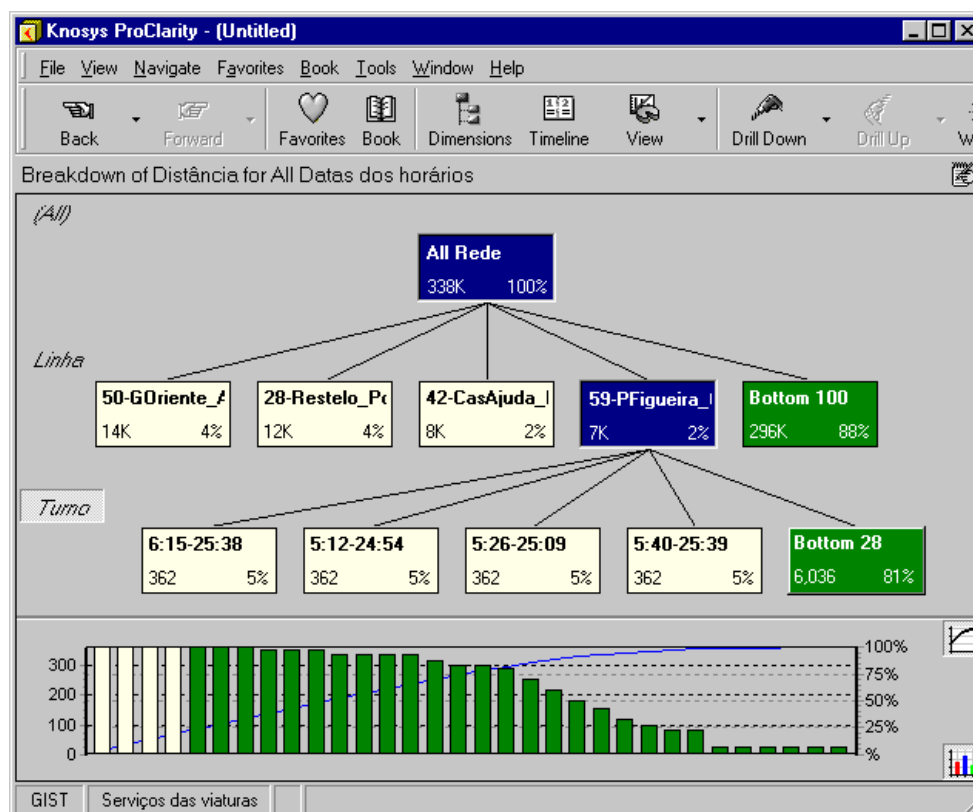
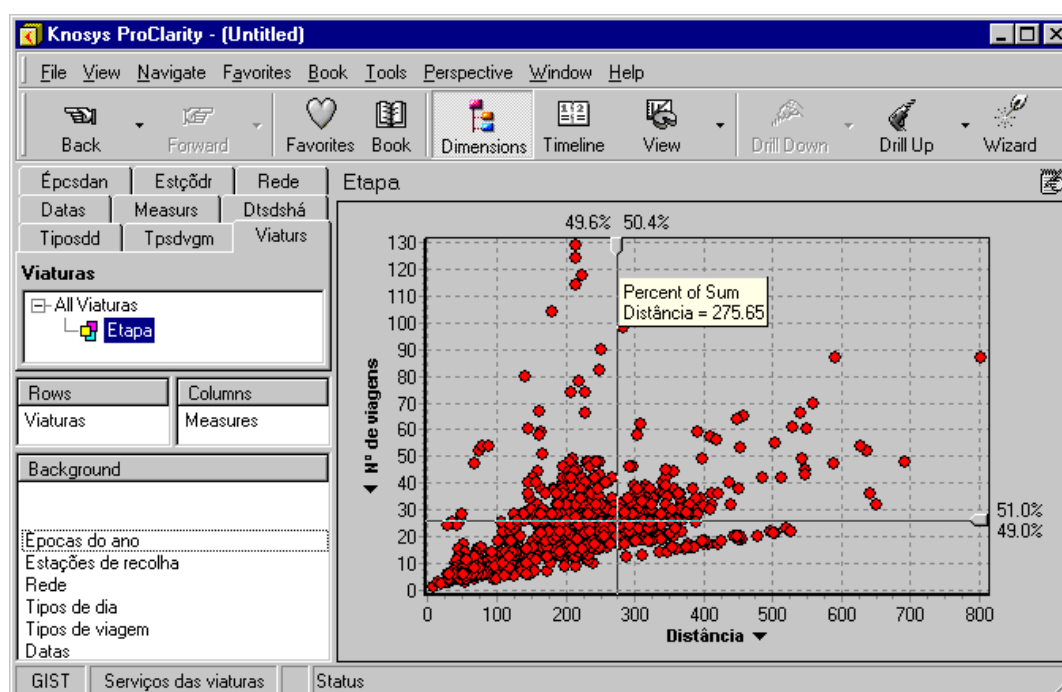
Na Figura 6.24 mostra-se a visualização em formato de gráfico de barras estando disponíveis outros tipos de gráficos.

Existem outros instrumentos de análise de grande utilidade. A decomposição em árvore permite detalhar um determinado indicador pelos critérios pretendidos. No exemplo apresentado na Figura 6.25 estuda-se a distância percorrida por todas as viaturas num dia – 338 mil quilómetros – discriminando-se este valor pelas linhas, ou seja, é possível saber quantos quilómetros são percorridos em cada linha num dia. Para a linha 59 discrimina-se quais os turnos que irão assegurar a distância referente a essa linha – 7 mil quilómetros. Ao todo serão 32 as viaturas que garantem o serviço nesta linha (as 4 visíveis mais as 28 referidas no rectângulo do canto inferior direito).

Name	Tempo Effect. Plan.	Tempo Effect. Real	Tempo Effect. Real. / Plan.
ANTONIO JOSE GOMES CASIMIRO	118.57	120.45	1.02
ANTONIO LOPES PEREIRA	136.49	161.00	1.18
ANTONIO LUCAS DA SILVA	195.08	218.14	1.12
ANTONIO MANUEL ALMEIDA RODRIGUES	168.17	179.15	1.07
ANTONIO MANUEL FERREIRA AGRELOS	157.10	142.30	0.91
ANTONIO MANUEL FERREIRA ALBUQUERQUE	194.28	195.15	1.00
ANTONIO MANUEL RIBEIRO	174.56	199.00	1.14
ANTONIO MANUEL VARELA ALVES	137.41	167.07	1.22
ANTONIO MARIA GODINHO ANTUNES	221.46	223.45	1.01
ANTONIO MARTINS PEREIRA	148.42	156.25	1.05
ANTONIO MENDES ANTÃO	142.08	142.00	1.00
ANTONIO MENDES FORMIGO	113.59	114.15	1.00
ANTONIO NUNES	108.02	112.17	1.04
ANTONIO NUNES CACHUCHO	140.22	145.00	1.03
ANTONIO NUNES JOAQUIM	188.39	231.00	1.23
ANTONIO NUNES PEREIRA	195.28	225.45	1.15
ANTONIO PEREIRA FERREIRA	181.45	189.51	1.04
ANTONIO PINTO CORREDOURA	75.45	75.45	1.00

Figura 6.23: *Knosys ProClarity 2.0* – matrizFigura 6.24: *Knosys ProClarity 2.0* – gráfico de barras

É ainda possível estudar a eventual existência de correlação entre dois indicadores. No exemplo apresentado na Figura 6.26 é possível verificar a existência de correlação positiva entre o número de viagens e a distância, referentes às etapas dos turnos. É ainda possível ver a mediana da distância que será de aproximadamente 276 quilómetros.

Figura 6.25: *Knosys ProClarity 2.0* – decomposição em árvoreFigura 6.26: *Knosys ProClarity 2.0* – análise de correlação

Para além dos exemplos apresentados pode-se recorrer à linguagem *Microsoft Visual Basic for Applications* ou à linguagem *MDX (MultiDimensional eXpressions)*. Esta última

linguagem pode ser vista como a equivalente à linguagem *SQL* para estruturas multi-dimensionais.

6.7 A etapa de implementação

Tal como foi referido na fase de desenho da arquitectura do Armazenamento de Dados é, agora, possível estimar a dimensão da Base de Dados de análise e controlo com mais rigor. Comparando os resultados obtidos na Tabela 6.13 com 50% dos valores apresentados na Tabela 6.10 (só se estão a comparar os valores referentes ao espaço ocupado pelos dados devendo-se, por isso mesmo, excluir as estimativas referentes aos índices) pode-se observar que apenas para o caso do Cubo ‘serviços das viaturas’ os valores são significativamente diferentes. Tal facto deve-se, entre outros factores, à poupança de espaço na tabela *trips* uma vez que a mesma viagem se repete por um período de tempo relativamente extenso, existindo para esse período apenas um registo na tabela *trips* (no cálculo efectuado assumiu-se a existência de dez datas de horários ao público, o que dá em média 36,5 dias em que uma mesma viagem se repete).

Tabelas	Nº de Registos	Nº de Colunas	Nº de bytes por registo nas colunas de tamanho fixo	Nº de colunas de tamanho variável	Nº máximo de bytes por registo nas colunas de tamanho variável	Espaço ocupado (Bytes)	Espaço ocupado (Mega Bytes)
AUX_COMPANIES	1	2	6	1	20	8.192	
AUX_NETWORK_TYPES	1	2	6	1	20	8.192	
AUX_PATH_TYPES	2	2	6	1	20	8.192	
NETWORK	350	7	30	2	40	32.768	
NETWORK_FACT	3.500	4	20	0		106.496	0,16
AUX_TRIPS_TYPES	2	2	6	1	20	8.192	
DAY_TYPES	3	2	6	1	20	8.192	
DEPOTS	7	2	6	1	20	8.192	
SHIFTS_STRETCHES	2.000	7	28	2	40	172.032	
TRIPS	35.000	9	44	1	20	2.760.704	
TRIPS_DAYS	12.775.000	2	10	0		245.669.888	
YEAR_SEASONS	3	2	6	1	20	8.192	237,12
CATEGORIES	13	2	6	1	20	8.192	
CREWS	2.500	3	10	1	20	114.688	
CREWS_DAILY_INFORMATION	836.945	14	84	0		79.724.544	
FREE_DAYS_GROUPS	10	2	6	1	20	8.192	
SITUATION_TYPES	100	2	6	1	20	8.192	76,16
Total:						328.663.040	313,44

Tabela 6.13: Espaço ocupado por um ano de dados (folha de cálculo explicada em [Microsoft Corporation, 1998])

Mesmo sabendo que os valores apresentados não contabilizam o espaço ocupado pelos Cubos mas apenas os dados da Base de Dados relacional, pode-se concluir com facilidade que o espaço em disco não é, de forma alguma, crítico. Uma vez que é aconselhável um servidor dedicado, dado o grande consumo de memória por parte do *SQL Server 7*, o espaço em disco habitualmente existente numa máquina *standard* (superior a 10 Giga

Bytes) é claramente suficiente. Quanto à memória RAM mínima para o sistema operativo do servidor e para os serviços *OLAP* do *SQL server 7* é de 64 *Mega Bytes* [Microsoft Corporation, 1998]. No entanto, é aconselhável um servidor com mais memória *RAM*, uma vez que o desempenho é fortemente dependente da memória *RAM* existente. Sugere-se, assim, um servidor com 256 *Mega Bytes* de memória *RAM* com possibilidade de expansão até 1 *Giga Byte*. Quanto aos clientes, a memória *RAM* mínima é de 32 *Mega Bytes* sendo aconselhável pelo menos 64 *Mega Bytes* com possibilidade de expansão. O espaço em disco duro necessário depende da ferramenta seleccionada para o cliente final.

Como já foi referido anteriormente, a etapa de implementação não foi realizada por se terem registado atrasos na implementação do sistema GIST98/EUROBUS nas empresas do consórcio GIST. No entanto, para uma melhor compreensão do método de desenvolvimento adoptado na construção do Módulo QIO, faz-se uma projecção das tarefas fundamentais a ter em consideração nesta etapa:

- Instalação dos equipamentos e *software* necessários.
- Configuração das ferramentas, nomeadamente a criação da Base de Dados, a criação de utilizadores e a definição da política de acessos.
- Preenchimento das tabelas *sys_dbs*, *sys_parameters* e *aux_*. O desempenho dos serviços *OLAP* e o espaço ocupado em disco devem ser monitorizados para assim ajustar a forma de armazenamento e as opções de agregação o melhor possível face à realidade existente.
- Definir a política de execução de cópias de segurança e de aquisição de dados adequada à realidade de cada empresa.
- Vistoriar com regularidade a tabela *sys_log* e fazer testes de evolução dos indicadores para a detecção de eventuais erros.

7 Conclusões

Para concluir avalia-se o trabalho desenvolvido e discutem-se algumas questões que se colocam naturalmente no final de um trabalho desta natureza, e que são:

- Quais os desenvolvimentos futuros para o Módulo QIO?
- Qual a melhor estratégia para o desenvolvimento de um Armazém de Dados?

7.1 Avaliação do trabalho desenvolvido

Este trabalho teve como principal objectivo a apresentação e discussão dos problemas associados ao desenvolvimento de sistemas de Armazenamento de Dados. Para tal, apresentaram-se os conceitos fundamentais e alguns métodos de desenvolvimento adequados para este tipo de sistemas. Esse trabalho foi concluído com a aplicação destes conceitos a um caso de estudo no âmbito do sistema GIST98/EUROBUS - um Sistema de Apoio à Decisão para o planeamento de empresas de transportes colectivos rodoviários de passageiros.

Um dos pontos fortes deste trabalho reside na síntese efectuada sobre os vários problemas associados ao desenvolvimento de sistemas de Armazenamento de Dados, dando uma boa visão de conjunto sobre este tema. A leitura desta dissertação será, com certeza, útil para quem se inicia nesta área de trabalho/estudo. Os resultados apresentados são, ainda, uma boa base de trabalho para quem pretender desenvolver um conjunto de indicadores para os agentes de decisão de empresas de transportes colectivos de passageiros, em particular, para a área de planeamento.

Os pontos mais fracos desta dissertação resultaram da falta de tempo para abordar todos os assuntos com a profundidade desejada. Assim, o estudo da dimensão da Base de Dados (subsecção 6.3.1) e o desenho do repositório dos Meta Dados (subsecção 6.5.3) poderiam ter sido mais aprofundados. Neste último caso veja-se o esquema proposto por [Jennings, 2000]. É também de notar a não concretização da etapa de implementação, por motivos já apresentados, o que impediu a realização de uma iteração completa do método de desenvolvimento.

7.2 Possíveis desenvolvimentos do Módulo QIO

A evolução do Módulo QIO pode ser equacionada através de duas perspectivas distintas. A perspectiva de uma posterior comercialização do sistema GIST98/EUROBUS ou a perspectiva de suprir as carências de informação por parte de cada uma das empresas do consórcio GIST.

7.2.1 Na perspectiva de evolução do sistema GIST 98/EUROBUS

Para que o Módulo QIO possa, de facto, cobrir toda a área de planeamento e controlo dos serviços das viaturas e dos tripulantes, fornecendo todos os indicadores pertinentes para esta área funcional, sugerem-se três aspectos a melhorar no actual sistema GIST98/EUROBUS:

- O escalonamento das viaturas, ou seja, um módulo ou conjunto de módulos semelhantes ao Módulos de Escalamento mas que tenham por objecto as viaturas em vez do pessoal tripulante, permitindo obter indicadores para o controlo do serviço das viaturas;
- A evolução da Base de Dados de informação ao público, nomeadamente através da obtenção de mais dados geográficos de forma a melhorar a informação sobre a rede comercial;
- A integração dos diversos módulos de forma a não perder informação oriunda de módulos a montante, por exemplo, garantir que a informação do serviço diário a realizar pelos tripulantes contém as viagens respectivas. Na actual versão do sistema GIST98/EUROBUS só é mantida, dos módulos anteriores, a informação estritamente necessária para a execução das tarefas pretendidas nos módulos a jusante. Esta opção, defensável do ponto de vista operacional, impede, objectivamente, a obtenção de indicadores da maior relevância. Usando o mesmo exemplo não é possível saber o tempo de condução, nem o planeado nem o realizado pelo motorista.

7.2.2 Na perspectiva das empresas do consórcio GIST

A forma como o Módulo QIO se enquadra na estratégia de informação das empresas do consórcio GIST pode ser tipificada da seguinte forma:

- Para empresas com intenção de construir um Armazém de Dados, o Módulo QIO foi, de certa forma, encarado como uma experiência piloto. Para a construção de um

Armazém de Dados, as soluções técnicas a adoptar podem não ser as mesmas daquelas que se utilizaram na construção do Módulo QIO, até porque a dimensão do problema será completamente diferente. No entanto, o envolvimento de várias pessoas das empresas nas várias fases do método de desenvolvimento, em particular na fase de estudo dos requisitos de negócio, assim como a obtenção dos primeiros resultados foram encarados como muito positivos por parte dessas empresas.

- Para empresas que não prevêem investir na construção de um Armazém de Dados, o Módulo QIO e o Módulo de Informação ao Público são encarados como os módulos que, de certa forma, completam o sistema GIST98/EUROBUS, já que integram a informação dos restantes módulos de forma orientada para fins bem definidos e da maior relevância para as empresas. No caso do primeiro módulo referido, o fim é a distribuição de informação para dentro da empresa, em particular aos agentes de decisão, enquanto no segundo caso o fim é a distribuição de informação para fora da empresa, ou seja, aos potenciais clientes.

7.3 Armazém de Dados ou Armazém de Dados Departamental?

A opção pelo desenvolvimento de vários Armazéns de Dados Departamentais independentes de forma parcelar e desfasada no tempo ou o desenvolvimento do Armazém de Dados como um todo são as alternativas a equacionar.

A opção de dividir o problema em vários e tratá-los de forma separada surge como uma forma de diminuir a sua complexidade diminuindo, simultaneamente, o tempo necessário para apresentação de resultados. Pretende-se, assim, resolver um dos factores mais críticos para o sucesso de um projecto de Armazenamento de Dados e que é a necessidade de gerir as expectativas dos clientes através da apresentação de resultados a curto prazo.

No entanto, segundo [Inmon, 1998], esta opção acarreta diversos riscos dos quais se destacam:

- Ao desenvolver isoladamente cada um dos Armazéns de Dados Departamentais impede-se a análise integrada da informação caso se adoptem critérios diferentes na definição da informação comum a vários departamentos ou funções.
- O grau de detalhe necessário a um departamento pode ser diferente daquele que é necessário para a empresa. Normalmente, os Armazéns de Dados têm informação mais

detalhada, enquanto os diversos departamentos só tem os resultados agregados que lhes são relevantes¹⁰.

Acresce a estes riscos o facto de se perder a oportunidade de racionalizar a comunicação entre os diversos departamentos da empresa em estudo e, ainda, o aumento do custo de manutenção do sistema, já que, com a eventual criação de outros Armazéns de Dados Departamentais independentes, esse custo será seguramente maior. A vantagem mais evidente reside na diminuição da dimensão do problema, tornando mais fácil a obtenção de resultados a curto prazo.

Em [Kimball et al, 1998] é defendida a definição das dimensões e dos indicadores garantindo que estejam em conformidade de departamento para departamento. Tal pressupõe a construção de uma matriz semelhante à apresentada na Tabela 6.5 mas que inclua não só as dimensões mas também os indicadores. Só após a realização deste trabalho é que se inicia a construção do Armazém de Dados começando por desenvolver a Base de Dados para o caso do departamento para o qual a equipa de trabalho se sinta mais apta e que necessite de dados oriundos de uma só Base de Dados fonte. Garante-se, assim, o desenvolvimento parcelar com as vantagens já referidas, assegurando que os dados são integrados (secção 2.1.2). Ou seja, a opção é a de construir um Armazém de Dados de forma parcelar com a garantia de que não se adoptam critérios diferentes na definição da informação comum a vários departamentos ou funções.

¹⁰ Esta consideração não é válida para todas as empresas, sendo no entanto aplicável a grande parte delas.

Anexo A: Indicadores propostos pelas empresas

STCP

Grupo Barraqueiro

Rodoviária de Lisboa

EVA Transportes

Horários do Funchal

Anexo B: Esquemas relacionais da Base de Dados de planeamento

Módulo Rede

Módulo Megalinhas

Módulo Viagens e Viaturas

Módulo Serviços

Módulo Administração

Módulo Gestão de Pessoal Tripulante

Módulo Pré-Escalamento

Módulos Geração de Escalas e Pós-Escalamento

Bibliografia e referências

- [Adelman & Oates, 1998] Sid Adelman, Joe Oates, “Data Warehouse Project Management”, Data Management Review, Maio 1998.
- [Barros et al, 1999] Teresa Cunha Barros, Dinis Mota Paes, João Falcão e Cunha, Gabriel David, “INFOBUS: Public Transport Information on the Word Wide Web”, apresentado em: “EMMSEC99 - European Multimedia, Microprocessor System and Electronic Commerce”, Estocolmo, Suécia, 1999.
- [Borges, 1994] José Luis Borges, “Sistema de apoio à geração de escalas de pessoal no planeamento operacional de sistemas de transportes colectivos”, Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, 1994.
- [Branco, 1989] Isabel M. Branco, “Algoritmos para modelos matemáticos de quasi-afecção e extensões”, Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa, 1989.
- [Corey et al, 1999] Michael Corey, Michael Abbey, Ian Abramson, Larry Barnes, Benjamin Taub, Rajan Venkitachalam, “SQL Server7 Data Warehousing”, Osborne/McGraw-Hill, 1999, pp. 126-127.
- [Couto, 1998] José Miguel Couto, “Uma heurística para apoio à geração de serviços de motoristas em transportes colectivos”, Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, 1998.
- [Date, 1990] C. J. Date, “An Introduction to Database Systems”, Vol 1 (5th Edition) Addison-Wesley, 1990, pp. 245-614.
- [Demarest, 1995] Marc Demarest, “A Data Warehouse Evaluation Model”, Oracle Technical Journal, Outubro 1995.
- [Dias, 1995] Teresa Galvão Dias, “Aplicação de algoritmos genéticos ao problema de geração de serviços de tripulações”, Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, 1995.
- [Ebel, 1998] Doug Ebel, “Data Warehousing: Start to Start”, NCR Corporation, 1998.

- [Ferreira, 1988] José Vasconcelos Ferreira, “Concepção de um módulo interactivo para geração de escalas de pessoal tripulante em transportes colectivos urbanos”, Tese de Mestrado, Instituto Superior Técnico da Universidade Técnica de Lisboa, 1988.
- [Ferreira et al, 1998] José Vasconcelos Ferreira, Teresa Galvão Dias, João Falcão e Cunha “A generic crew rostering DSS”, apresentado em: “EURO XVI – 16th European Conference on Operational Research”, Bruxelas, Bélgica, 1998.
- [Gardner, 1998] Stephen R. Gardner, “Building the Data Warehouse”, Communications of the ACM, Setembro 1998.
- [Hackney, 1997] Douglas Hackney, “Picking a Data Mart Tool”, Data Management Review, Outubro 1997.
- [Hadden & Company, 1999] Hadden & Company, <http://www.hadden-kelly.com>, leitura efectuada em Dezembro de 1999.
- [Hufford, 1996] Duane Hufford, “Data Warehouse Quality”, Data Management Review, Janeiro 1996.
- [ICAT & INEGI, 1997] ICAT, INEGI, “Especificação do Sistema GIST98”, documento interno, 1997.
- [INE, 1998] Instituto Nacional de Estatística, Estatísticas do transporte rodoviário de passageiros e de mercadorias, 1998.
- [Inmon, 1996] W. H. Inmon, “Building the Data Warehouse”, Wiley, 1996.
- [Inmon, 1998] W. H. Inmon, “Data Mart ≠ Data Warehouse”, Data Management Review, Maio 1998.
- [Jennings, 2000] Mike Jennings, “The Generic Meta Data Repository”, Data Management Review, Outubro 2000.
- [Kelly, 1996] Sean Kelly, “Data Warehousing – the route to mass customization”, Wiley, 1996, pp. 55.
- [Kimball et al, 1998] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, “The Data Warehouse Lifecycle Toolkit”, Wiley, 1998, pp. 137-314.

- [Lee, 1998] Gloria Lee, “Successful Data Warehouses: Lessons From the Front Line”, Data Management Review, Fevereiro 1998.
- [Mesquita & Paixão, 1990] Marta Mesquita, José Pinto Paixão, “Multi depot vehicle scheduling problems: a new heuristic based on quasi-assignment algorithms”, apresentado em: “5th international workshop on computer-aided scheduling of public transport”, Montreal, Canadá, 1990.
- [Microsoft Corporation, 1998] Microsoft Corporation, “SQL Server Books Online”, livros em formato digital que fazem parte integrante do *SQL Server 7*, 1998.
- [Nunes & Cunha, 1998] Nuno Jardim Nunes, João Falcão e Cunha, “Case Study: SITINA - A Software Engineering Project Using Evolutionary Prototyping”, apresentado em: “*Proc. EMMSAD’98, the 3rd CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design*, Keng Siau (ed.)” em conjunto com “*CAiSE’98 the 10th Conference on Advanced Information Systems Engineering*”, Pisa, Itália, 1998.06.8-9, pp. Q 1-12
- [OLAP Council, 1995] OLAP Council, “OLAP and OLAP Server definitions”, <http://www.olapcouncil.org/research/glossaryly.htm>, Janeiro 1995.
- [Onder & Nash, 1998] John Onder, Todd Nash, “Building a Business-Driven Data Warehouse”, Data Management Review, Outubro 1998.
- [Paes, 1999] Dinis Paes, “Estudo da aplicação do CASE Oracle Designer/2000 ao projecto GIST”, Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, 1999, pp. 17.
- [Paixão, 1987] José Pinto Paixão, “A quasi-assignment algorithm for the bus scheduling problem”, Networks n°17, 1987, pp. 249-269.
- [Paixão & Pato, 1989] José Pinto Paixão, Margarida Vaz Pato, “A structural lagrangean relaxation for two duty period bus driver scheduling problem”, European Journal of Operational Research n°39, 1989, pp. 213-222.
- [Paul, 1997] Lauren Gibbons Paul, “Anatomy of a Failure”, CIO Enterprise Magazine, Novembro 1997.

- [Pedrosa et al, 2000] Dulce Pedrosa, Fernando Vieira, Luís Pedro Ferreira, José Vasconcelos Ferreira, “Manual do Utilizador - Tutorial dos Módulos de Escalamento”, manual do sistema GIST98/EUROBUS, versão 2.0.0, 2000.
- [Pendse & Creeth, 1997] Nigel Pendse, Richard Creeth, “The OLAP Report”, Business Intelligence Limited, 1997.
- [Pendse & Creeth, 2000a] Nigel Pendse, Richard Creeth, “The OLAP Report”, <http://www.olapreport.com>, leitura efectuada em Agosto de 2000.
- [Pendse & Creeth, 2000b] Nigel Pendse, Richard Creeth, “The OLAP Report”, <http://www.olapreport.com>, leitura efectuada em Outubro de 2000.
- [Perkins, 1998] Alan Perkins, “Critical Success Factors for Data Warehouse Engineering”, Visible Systems Corporation, 1998.
- [Portugal, 1998] Rita Portugal, “Metaheurísticas para a geração de serviços para o pessoal tripulante”, Tese de Mestrado, Faculdade de Ciências da Universidade de Lisboa, 1998.
- [Raden, 1995/6] Neil Raden, “Data Modeling – Star Schema 101”, <http://www.archer-decision.com>, 1995/6.
- [Ranito, 1999] João Ranito, acetatos de um seminário realizado no âmbito da cadeira de Sistemas de Informação e Bases de Dados da Licenciatura em Engenharia Informática e Computadores da Faculdade de Engenharia da Universidade do Porto, 1999.
- [Sousa et al, 2000] Jorge Pinho de Sousa, João Falcão e Cunha, Rui Campos Guimarães, José Pinto Paixão, “GIST - Um Sistema de Apoio à Decisão para o Planeamento Operacional de Transportes Colectivos”, artigo que faz parte integrante do livro: “Casos de Aplicação da Investigação Operacional”, Carlos Henggeler Antunes e Luís Valadares Tavares, McGraw-Hill Portugal, 2000, pp. 108-130.
- [STCP, 1998] STCP, “DEP – Indicadores de gestão”, documento interno, Janeiro 1998.
- [The Data Warehousing Institute, 1998] The Data Warehousing Institute, 1998 Data Warehousing Buyers Guide, <http://www.dw-institute.com/buyersguide/>, 1998.
- [Thomsen, 1997] Erik Thomsen, “OLAP Solution – Building Multidimensional Information Systems”, Wiley, 1997, pp. 483-489.

- [Turban & Aronson, 1998] Efraim Turban, Jay E. Aronson, “Decision Support Systems and Intelligent Systems”, Prentice Hall, 1998, pp. 73-78.
- [Watterson, 1997] Karen Watterson, “Attention, Data-Mart Shoppers”, BYTE Magazine, Julho 1997.
- [Welch, 1998] J. D. Welch, http://www.datawing.com/BOOKS_F.HTM, 1998.